Market Quality and Price Impact of High-Frequency Trading and its Regulation

Dissertation

submitted to the

Faculty of Business, Economics and Informatics

of the University of Zurich

to obtain the degree of

Doktor der Wirtschaftswissenschaften, Dr. oec.

(corresponds to **Doctor of Philosophy, PhD**)

presented by

Jakub Rojček

from Slovakia

approved in September 2016 at the request of

Prof. Dr. Thorsten Hens

Prof. Dr. Ramazan Gençay

Prof. Dr. Felix Kübler

The Faculty of Business, Economics and Informatics of the University of Zurich hereby authorizes the printing of this dissertation, without indicating an opinion on the views expressed in the work.

Zurich, September 21, 2016.

The Chairman of the Doctoral Board: Prof. Dr. Steven Ongena.

Acknowledgements

"For a true writer, each book should be a new beginning where he tries again for something that is beyond attainment. He should always try for something that has never been done or that others have tried and failed. Then sometimes, with great luck, he will succeed."

— *Ernest Hemingway*

I dedicate this thesis to my beloved family.

First and foremost, I want to express my deepest gratitude to Prof. Dr. Thorsten Hens and PD Dr. Alexandre Ziegler for their joint work of supervising my Ph.D. thesis, for their continuous support, patient encouragement and crystal guidance. I want to especially thank them for giving me the chance to undergo the rewarding transition to becoming a researcher.

Second, I would like to thank Prof. Dr. Ramo Gençay from Simon Fraser University for sparking my interest in high-frequency trading and for giving me the opportunity to collaborate on a project during an extended, invigorating research stay at the department of Economics at SFU in Vancouver, Canada. Here, I could meet and collaborate with Ramo's former like-minded Ph.D. students Soheil Mahmoodzadeh and Michael Tseng, from whom I have learned immensely during our intense reading group.

I am very grateful to Prof. Dr. Felix Kübler from Swiss Finance Institute (SFI) and University of Zurich for generously accepting to be on my Ph.D. thesis committee.

Next, I am deeply indebted to Prof. Dr. Michel Habib and Prof. Dr. Jean-Charles Rochet from the University of Zurich for their teaching and financial support in the second phase of my Ph.D. studies. I am truly thankful for their kindness and helpfulness. On top of that, I am especially thankful to Michel for his counsel on numerous occasions.

I would like to express my gratitude to Prof. Dr. Angelo Ranaldo from the University of Sankt Gallen for his invaluable feedback on my research and sharing his wide knowledge of market microstructure in a doctoral course. During my Ph.D. studies I have also received valuable comments from Dr. Stefano Balietti, Christopher Hemmens, Prof. Dr. Michel Habib, Prof. Dr. Joel Hasbrouck, Prof. Dr. Harald Hau, Prof. Dr. Boyan Jovanovic, Prof. Dr. Hubert Kempf, Prof. Dr. Albert Menkveld, Prof. Dr. Artem Neklyudov, Prof. Dr. Per Östberg, Prof. Dr. Marc Paolella, Prof. Dr. Jean-Charles Rochet and Prof. Dr. Alexander Wagner, to all of whom I am thankful.

The faculty members and the Ph.D. students at the Department of Banking and Finance at University of Zurich deserve a special thanks for their valuable feedback on my research and for creating a stimulating environment, be it by means of organized seminar discussions, random meetings in the corridors, or in other social events. It was my pleasure, to share the premises and ideas with Chris Bardgett, Nilüfer Caliskan, Pascal Caversaccio, Sabine Elmiger, Felix Fattinger, Fulvia Fringuellotti, Nicoletta Gabrielli, Marco Gambacciani, Elise Gourier, Ethem Ibrahim Güney, Manish Gupta, Regina Hammerschmid, Robert Huitema, Meriton Ibraimi, Manuel Kannenberg, Kerstin Kehrle, Jochen Krause, Dominika Kryczka, Felix Matthys, Elisabeth Megally, Kevin Meyer, Cosimo Munari, Magnus Nyboe, Diego Ostinelli, Anca Pana, Ivan Petzev, Pawel Polak, Tatjana-Xenia Puhan, Thomas Richter, Cornelia Rösler, Kim Schartz, Simon Scheidegger, Mario Šikić, Anastasiia Sokko, Felix Stang, Lujing Su, Adriano Tosi, Bruno Troja, Nikola Vasiljević, Patrick Walker, Zexi Wang, Jiri Woschitz, and Hanlin Yang. I should not forget my fellow Ph.D. students from other corners of the world: Vincent Bogousslavsky, Ilya Dergunov, Edouard Djeutem, Eloy Fisher, Thomas Geelen, Nataliya Gerasimova, Sharon Greenblum, Martina Jašová, Amir Khalilzadeh, Xiaowen Lei, Carolina Mattsson, Veridiana De Andrade Nogueira, Matthias Rupprecht, Jovan Stojković, Davide Tedeschini, Hanh Tong Thi, and Christopher Verzijl, all of whom I met at various research events, workshops, Ph.D. courses and research visits. I am grateful to all of them for their comments, suggestions regarding my research, brief collaborations, interesting discussions, and friendship. I would also like to thank Marilyn Barja, Martine Baumgartner, Susanne Erber, Eckart Jäger, Mira Jovanović, Sergio Maffioletii, Björn Maurer, Beatrice Rothenbach-Seiler, Michelle Rüegg, Franziska Spycher, and Stefan Widmer for helping me to navigate the administrative, organizational, and technical matters.

Last but not least, I wish to thank my family for their endless love and unconditional support. I would like to express my deepest gratitude and admiration to my wife and life partner Eva for synchronising our internal clocks during the ups and downs that the Ph.D. studies brought about.

Zürich, 30 May 2016 Jakub Rojček

Contents

List of Figures

List of Tables

Chapter 1

Abstracts

High-Frequency Trading in Limit Order Markets: Equilibrium Impact and Regulation

Jakub Rojček and Alexandre Ziegler

Abstract

We investigate the impact of high-frequency trading (HFT) on market quality and investor welfare using a dynamic general limit order book model. We find that while the presence of HFT always improves market quality under symmetric information, under asymmetric information this is the case only if competition between high-frequency traders is sufficiently strong. While HFT does not negatively impact investor welfare, it reduces the welfare of slow speculators. The flexibility of the model allows investigating the effect of the main recent regulatory initiatives designed to curb HFT on market quality and investor welfare. We consider minimum resting time rules, cancellation fees, transaction taxes, rebate fee structures, and speed bumps. While some of these regulations lead to improvements in a number of market quality measures, this generally does not translate into higher welfare for long-term investors. Rather, the main effect of such regulations is to generate wealth transfers from high-frequency traders to slow speculators. These regulations therefore appear inadequate to enhance investor welfare in the presence of HFTs. Of the different measures, transaction taxes are the least harmful; while they reduce welfare roughly by the amount of the tax, they do not significantly worsen market quality. The common practice by exchanges of granting rebates to limit orders is detrimental to market quality and investor welfare, causing both higher effective spreads and longer execution times.

Keywords: High-frequency trading, Regulation, Market quality

JEL classification: G14 · G28 · C63 · C73 · D82

Price Impact of Aggressive Liquidity Provision

Ramazan Gençay, Soheil Mahmoodzadeh, Jakub Rojček and Michael C. Tseng

Abstract

This paper analyzes brief episodes of high-intensity quotes turnover and revision—"bursts" in quotes—in the U.S. equity market. Such events occur very frequently, around 400 times a day for actively traded stocks. We find significant price impact associated with this market-maker initiated event, about five times higher than during non-burst periods. Bursts in quotes are concurrent with short-lived structural breaks in the informational relationship between market makers and market takers. During bursts, market makers no longer passively impound information from order flow into quotes—a departure from the traditional market microstructure paradigm. Rather, market makers significantly impact prices during bursts in quotes. Further analysis shows that there is asymmetry in adverse selection between the bid and ask sides of the limit order book and only a sub-population of market makers enjoys an informational advantage during bursts. Our results call attention to the need for a new microstructure perspective in understanding modern high-frequency limit order book markets.

Keywords: Price Impact, Burst, High-Frequency Trading, Market Quality, Adverse Selection

JEL classification: G14 · G28 · C58.

A Model of Price Impact and Market Maker Latency

Jakub Rojček

Abstract

Price impact measures the difference between the best quoted price and the realized price as a function of order size. This paper analyzes how price impact depends on the latency that a market maker is subject to. I propose a tractable model which allows incorporating both order size and latency effects as determinants of price impact. The model is solved analytically and is novel in the theoretical microstructure literature. Larger latency increases adverse selection costs to the market maker and reduces his probability of trading with a slow investor. A larger order size decreases the slow trader's outside option, making him susceptible to accept a worse price for his trade. It is shown that the first-order effect of increased latency and increased order size is to increase price impact. Their joint impact is also positive. When the probability of trading is taken into consideration, the utility of the slow institutional investor decreases with latency.

Keywords: Price Impact, High-Frequency Trading, Trade Size, Latency, Market Quality, Welfare

JEL classification: D53 · G14 · C72.

Hochfrequenzhandel in Limit-Order-Märkten: Auswirkungen auf das Marktgleichgewicht und Regulierung

Jakub Rojček and Alexandre Ziegler

Zusammenfassung

Wir untersuchen die Auswirkungen des Hochfrequenzhandels (HFT) auf die Marktqualität und die Wohlfahrt der Investoren anhand eines allgemeinen dynamischen Limit-Orderbuch-Modells. Wir finden, dass die Anwesenheit von HFT unter symmetrischer Information die Marktqualität immer verbessert. Unter asymmetrischer Information ist dies jedoch nur dann der Fall, wenn der Wettbewerb zwischen Hochfrequenzhändlern ausreichend stark ist. Während HFT das Wohlergehen der Investoren nicht negativ beeinflusst, reduziert es das Wohlergehen langsamer Spekulanten. Die Flexibilität des Modells ermöglicht es, die Auswirkungen der wichtigsten jüngsten Regulierungsansätze zu untersuchen, welche erwogen wurden, um Hochfrequenzhandel zu begrenzen. Wir betrachten Mindestruhezeitregeln, Stornogebühren, Transaktionssteuern, Rabattstrukturen und sogenannte Bodenschwellen (Speed Bumps). Während einige dieser Regelungen zu Verbesserungen in einer Reihe von Marktqualitätsmassen führen, ist dies in der Regel nicht mit höherer Wohlfahrt für langfristige Investoren verbunden. Vielmehr besteht der Haupteffekt dieser Regulierungen in einer Vermögensübertragung von Hochfrequenzhändlern zu langsamen Spekulanten. Diese Regelungen sind daher ungeeignet, um die Wohlfahrt der Anleger beim Vorliegen von HFT zu erhöhen. Von den verschiedenen Massnahmen sind Transaktionssteuern am wenigsten schädlich. Während sie die Wohlfahrt ungefähr um den Betrag der Steuer reduzieren, verschlechtern sie die Marktqualität nicht wesentlich. Die gängige Praxis vieler Börsen, Rabatte für Limit-Aufträge zu gewähren, reduziert die Marktqualität und die Wohlfahrt der Anleger, indem sie höhere effektive Geld-Brief-Spannen und längere Ausführungszeiten verursacht.

Schlüsselwörter: Hochfrequenzhandel, Regulierung, Marktqualität

JEL Klassifizierung: G14 · G28 · C63 · C73 · D82

Die Auswirkung agressiver Liquiditätsbereitstellung auf Preise

Ramazan Gençay, Soheil Mahmoodzadeh, Jakub Rojček and Michael C. Tseng

Zusammenfassung

Dieser Artikel untersucht kurze Perioden, in welchen Gebote mit einer hohen Intensität auf dem US-Aktienmarkt abgegeben und revidiert werden – sogenannte "Bursts". Solche Ereignisse treten sehr häufig auf, etwa 400 Mal pro Tag für aktiv gehandelte Aktien. Diese durch Market-Maker initiierte Ereignisse haben erhebliche Auwirkungen auf Preise, etwa fünfmal stärker als in anderen Perioden. Zusätzlich sind diese Perioden mit kurzlebigen strukturellen Brüchen in der Informationsbeziehung zwischen Market-Maker und Market-Taker verbunden. Während Burst-Perioden lassen die Market-Maker Information aus dem Auftragsablauf nicht mehr passiv in ihre Gebote einfliessen, sondern beeinflussen die Preise aktiv — was eine Abkehr vom traditionellen Paradigma der Markt-Mikrostruktur bedeutet. Nähere Analysen zeigen, dass eine Asymmetrie in der adversen Selektion zwischen der Kauf- und der Verkaufs-Seite des Limit-Orderbuches besteht und nur ein Teil der Market-Maker einen Informationsvorteil während Bursts geniesst. Unsere Ergebnisse zeigen die Notwendigkeit einer neuen Mikrostruktur-Perspektive auf, um moderne Hochfrequenz-Limit-Orderbuch-Märkte zu verstehen.

Schlüsselwörter: Preisauswirkung, Burst, Hochfrequenzhandel, Marktqualität, Adverse Selektion

JEL Klassifizierung: G14 · G28 · C58.

Ein Modell von Preisauswirkung und Market Maker-Latenz

Jakub Rojček

Zusammenfassung

Preisauswirkung misst die Differenz zwischen dem besten angebotenen Preis und dem realisierten Preis als Funktion der Auftragsgrösse. Dieser Artikel untersucht, wie die Preisauswirkung von der Latenz abhängt, der ein Market Maker unterliegt. Es wird ein analytisches Modell vorgestellt, das sowohl die Auftragsgrösse als auch die Latenz als Determinanten der Preisauswirkung erfasst. Das Modell wird analytisch gelöst und ist in der theoretischen Mikrostrukturliteratur neuartig. Eine grössere Latenz erhöht die Kosten der adversen Selektion für den Market Maker und reduziert die Wahrscheinlichkeit, dass er mit langsamen Investoren handelt. Eine grössere Auftragsgrösse verschlechtert die Alternativen des langsamen Händlers und veranlasst ihn, einen schlechteren Transaktionspreis zu akzeptieren. Es wird gezeigt, dass Latenz und Auftragsgrösse die Preisauswirkung erhöhen. Ihre gemeinsame Wirkung ist ebenfalls positiv. Wird die Handelswahrscheinlichkeit berücksichtigt, so reduziert Latenz den Nutzen langsamer institutioneller Anleger.

Schlüsselwörter: Preisauswirkung, Hochfrequenzhandel, Auftragsgrösse, Latenz, Marktqualität, Wohlfahrt

JEL Klassifizierung: D53 · G14 · C72.

Part I

Introduction

Chapter 2

Introduction and Summary of Research Results

High-frequency trading emerged as a set of algorithmic trading strategies that focus on the ability to react to new information faster than anybody else, or more precisely, faster than any other machine. The relative speed advantage combined with the ability to process information allowed traders with otherwise no intrinsic motives to trade to earn significant profits. The quest for a speed advantage led trading companies to explore the absolute speed limits by investing immense sums in technological, software, and communication development. The rise of trading machines competing on speed raised questions of great importance to our society, whose economy in its core relies on an effective and efficient allocation of capital. Is high-frequency trading good or bad for market quality? Does it make spreads smaller and help narrow the distance between assets' fundamental value and their price faster? Are execution costs higher because machines learn fundamental traders' trading intentions better? Does the price impact of a large trade change with higher speed? Do unpredictable jumps in prices emerge as an outcome of relative speed competition? Does the ability to change quotes at striking speed lead to price manipulation? Should high-frequency trading be regulated, and if so, how? Would that lead to welfare transfers among different groups of traders? How can we make the markets serve their purpose better?

This doctoral thesis entitled "Market Quality and Price Impact of High-frequency Trading and Its Regulation", aims at answering some of the above questions. It consists of three different papers, which utilize different methodologies in order to answer a subset of the above questions.

The first paper presented in Chapter 3, named "High-Frequency Trading in Limit Order Markets: Equilibrium Impact and Regulation", investigates the impact of high-frequency trading on market quality and investor welfare using a dynamic general limit order book model. This is joint work with Alexandre Ziegler. The paper finds that while the presence of HFT always improves market quality under symmetric information, under asymmetric information this is the case only if competition between high-frequency traders is sufficiently strong. While HFT does not negatively impact investor welfare, it reduces the welfare of slow speculators. The flexibility of the model allows investigating the effect of the main recent regulatory initiatives designed to curb HFT on market quality and investor welfare. The paper considers minimum resting time rules, cancellation fees, transaction taxes, rebate fee structures, and speed bumps. While some of these regulations lead to improvements in a number of market quality measures, this generally does not translate into higher welfare for long-term investors. Rather, the main effect of such regulations is to generate wealth transfers from high-frequency traders to slow speculators. These regulations therefore appear inadequate to enhance investor welfare in the presence of HFTs. Of the different measures, transaction taxes are the least harmful; while they reduce welfare roughly by the amount of the tax, they do not significantly worsen market quality. The common practice by exchanges of granting rebates to limit orders is detrimental to market quality and investor welfare, causing both higher effective spreads and longer execution times.

The second paper presented in Chapter 4 is an empirical study and is entitled "Price Impact of Aggressive Liquidity Provision." It is joint work with Ramo Gençay, Soheil Mahmoodzadeh, and Michael Tseng. In this paper, we analyze brief episodes of high-intensity quotes turnover and revision—"bursts" in quotes—in the U.S. equity market. Such events occur very frequently, around 400 times a day for actively traded stocks. We find significant price impact associated with this market-maker initiated event, about five times higher than during non-burst periods. Bursts in quotes are concurrent with short-lived structural breaks in the informational relationship between market makers and market takers. During bursts, market makers no longer passively impound information from order flow into quotes—a departure from the traditional market microstructure paradigm. Rather, market makers significantly impact prices during bursts in quotes. Further analysis shows that there is asymmetry in adverse selection between the bid and ask sides of the limit order book and only a sub-population of market makers enjoys an informational advantage during bursts. The paper's results call attention to the need for a new microstructure perspective in understanding modern high-frequency limit order book markets.

The last paper of this dissertation, "A Model of Price Impact and Market Maker Latency", is presented in Chapter 5. It provides an analytical model of price impact as a function of latency, which is the delay traders suffer and which can be thought of as the inverse of trading speed. Price impact measures the difference between the best quoted price and the realized price as a function of order size. The paper analyzes how price impact depends on the latency that a market maker is subject to. It presents a tractable model which allows incorporating both order size and latency effects as determinants of price impact. The model is solved analytically and is novel in the theoretical microstructure literature. Larger latency increases adverse selection cost to the market maker and reduces the probability of trading with a slow investor. A larger order size decreases the slow trader's outside option, making him susceptible to accept a worse price for his trade. It is shown that the first-order effect of increased latency and increased order size is to increase price impact. Their joint impact is also positive. When the probability of trading is taken into consideration, the utility of slow institutional investors decreases with latency.

Each paper of the dissertation is followed by the related tables and figures. Proofs and supplementary theoretical results are given in the appendix. The cumulative bibliography of all three research papers appears at the end of the thesis.

Part II

Research Papers

Chapter 3

High-Frequency Trading in Limit Order Markets: Equilibrium Impact and Regulation

Jakub Rojček and Alexandre Ziegler[1]

Abstract

We investigate the impact of high-frequency trading (HFT) on market quality and investor welfare using a dynamic general limit order book model. We find that while the presence of HFT always improves market quality under symmetric information, under asymmetric information this is the case only if competition between high-frequency traders is sufficiently strong. While HFT does not negatively impact investor welfare, it reduces the welfare of slow speculators. The flexibility of the model allows investigating the effect of the main recent regulatory initiatives designed to curb HFT on market quality and investor welfare. We consider minimum resting time rules, cancellation fees, transaction taxes, rebate fee structures, and speed bumps. While some of these regulations lead to improvements in a number of market quality measures, this generally does not translate into higher welfare for long-term investors. Rather, the main effect of such regulations is to generate wealth transfers from high-frequency traders to slow speculators. These regulations therefore appear inadequate to enhance investor welfare in the presence of HFTs. Of the different measures, transaction taxes are the least harmful; while they reduce welfare roughly by the amount of the tax, they do not significantly worsen

[1]We are grateful to Ronald Goettler for sharing his code. We have benefited from the comments of Ramo Gençay, Harald Hau, Christopher Hemmens, Soheil Mahmoodzadeh, Angelo Ranaldo, Michael Tseng, and of participants at the Multi-Agent Simulation and Global Issues workshop at the University of Tokyo, the Swiss Finance Institute Research Days in Gerzensee, and the University of Zurich. All errors and omissions remain ours.

market quality. The common practice by exchanges of granting rebates to limit orders is detrimental to market quality and investor welfare, causing both higher effective spreads and longer execution times.

Keywords: High-frequency trading, Regulation, Market quality

JEL classification: G14 · G28 · C63 · C73 · D82

3.1. Introduction

The U.S. Securities and Exchange Commission (SEC) describes high-frequency trading (HFT) as "one of the most significant market structure developments in recent years" (SEC, 2010, p. 45). High-frequency traders (HFTs) nowadays account for over half of the volume on many stock, futures and options exchanges. After the Flash crash that occurred in May 2010, a controversy around the impact of HFT on market quality arose in the public discussion, the academic literature, and among regulators and exchanges. Of particular concern are the lack of knowledge about HFTs' strategies, their potentially destabilizing effect on markets in periods of turmoil, and the large volume of quotes not leading to trades which HFTs submit to exchanges, so-called high-frequency spam.[2]

Concerns about the impact of HFT on the functioning of markets are not limited to the U.S. For example, a new European Directive (so-called MiFID 2, 2011) identifies "specific regulatory and supervisory measures necessary in order to adequately deal with the potential threats for the orderly functioning of markets arising from algorithmic and high-frequency trading." HFT and make/take fee structures – the practice common among exchanges of granting fee rebates to qualified market participants providing liquidity – are also among concerns listed in a consultation paper by the Committee of European Securities Regulators (2010). While regulators have been mostly concerned about the impact of HFT on market quality and the welfare of long-term investors, exchanges have been mostly worried about high-frequency spam, as the practice has raised their infrastructure costs without leading to a commensurate increase in trading volume.

The goal of this paper is to assess the impact of HFT on market quality and investor welfare and to investigate the suitability of a number of regulatory measures that have been suggested to improve market quality in the presence of HFT. These measures include a *minimum resting time* for quotes, the

[2] The term *high-frequency spam* refers to the practice of rapidly canceling and changing quotes without even changing price and with no intention of executing, i.e. anticipatory strategies used to reveal liquidity and trade ahead of it (Hunsader, 2011).

imposition of *cancellation fees* when the ratio of quotes to trades exceeds a certain value, *transaction taxes*, so-called *speed bumps* enforcing minimum delays between the time an order is submitted and the time it hits the order book, and the use of *make-take fees* (or *rebate fee structures*) to provide HFTs incentives to make markets.

In spite of the fact that little is known about the impact of these measures on market quality and investor welfare, a number of them have been introduced in several countries either alone or in combination:

1. *Minimum resting time* or similar rules have been introduced in Europe and the U.S. The Eurex options exchange imposes a minimum resting time for quotes as part of market maker obligations. The European Parliament has enacted a requirement for HFTs to maintain their quotes for at least half a second, and Italy has introduced a tax of 0.02% on orders that are cancelled within half a second. In the U.S., minimum resting time requirements have been derived from FINRA rule 5210, which requires quotes to represent bona fide intention to trade. This rule has been explicitly adopted by the NYSE and NYSE Arca in 2011 (see SEC 2011a, 2011b), and the CME enacted a similar regulation in its rule 575 in 2014 (see CFTC, 2014).

2. *Cancellation fees* have been introduced by the Canadian regulatory authority IIROC with the explicit aim of curbing high-frequency spam. The Milan Stock Exchange and the NASDAQ Stock Market introduced an excess order fee for high quote-to-trade ratios in April and July 2012 (see SEC, 2012). Cancellation fees are also imposed by Eurex for order-to-trade ratios exceeding 5, on NASDAQ OMX for ratios exceeding 100, as well as on the German stock exchange. The new MiFID2 European Directive also foresees the introduction of cancellation fees to curb excessive quote-to-trade ratios.

3. *Transaction taxes* are currently imposed in roughly 40 countries globally and mostly serve to finance regulatory activities related to the operation of financial markets. The U.S. imposes a financial transaction tax of 0.0034%. The U.K. collects a stamp duty of 0.5%, but many intermediary transactions are exempted. France introduced a financial transaction tax of 0.2% in August 2012, and Italy a tax of 0.12% in 2013. In January 2013, 11 European countries approved the principle of a financial transaction tax, but postponed its implementation until mid-2016.

4. *Speed bumps* have been introduced by NASDAQ OMX and by a newly founded exchange, IEX, in October 2013. At IEX, the speed bump takes the form of a minimal physical distance between clients' servers and the exchange's main server and is imposed on all orders. NASDAQ OMX applies the speed bump to market orders only, presumably to encourage liquidity provision.

5. *Make-take fees* giving incentives to liquidity providers are used by numerous exchanges in North America (NASDAQ, NYSA Arca, BATS Exchange, NSX, BATS Options Exchange, Boston Options Exchange, International Securities Exchange, TSX) and Europe (Euronext, OMX exchanges, Chi-X). Many other exchanges, including the London Stock Exchange, XETRA and SIX Swiss Exchange grant rebates to designated market makers or other significant providers of liquidity.

We assess the impact of HFT and of proposed regulations on market quality and investor welfare using a unified framework based on the theoretical model of Goettler, Parlour and Rajan (2005, 2009). Goettler et al. develop a theoretical model of a dynamic limit order market with strategic players and asymmetric information and solve for Markov perfect equilibria numerically using an extension of the simulation-based algorithm of Pakes and McGuire (2001) for complete information games. We conduct our investigation by comparing the properties of market equilibria in different settings, specifically settings with HFTs to settings without HFTs, and settings where specific regulatory proposals are implemented to settings where they are not. In addition to measures of market quality and overall welfare, we investigate the impact of HFT and proposed regulations on specific groups of market participants, namely long-term investors and speculators.[3]

We find that the presence of HFTs improves most common measures of market quality. Specifically, markets with HFTs exhibit narrower spreads, improved price discovery and higher depth. These results hold without qualification under symmetric information. When HFTs have an informational advantage over slow traders, however, improvements in market quality arise only if competition between them is sufficiently strong. We also find that HFT does not negatively impact investor welfare. However, it leads to a reduction in the welfare of slow speculators, who are crowded out by HFTs.

We also find that none of the proposed regulations consistently improves the welfare of long-term investors. Admittedly, some of the proposed regulations do lead to an improvement in a number

[3] Consistent with the terminology in Goettler et al. (2005, 2009), we use the term speculator to denote a trader with zero private valuation for an asset; hence the term includes HFTs.

of market quality measures. For example, minimum resting time rules reduce trading costs under asymmetric information, while speed bumps increase book depth and decrease spreads. However, none of the improvements in market quality translates into higher welfare for long-term investors. Rather, the main effect of these regulations is to generate wealth transfers from high-frequency traders to slow speculators. These regulations therefore appear inadequate to enhance investor welfare in the presence of HFTs. Of the different measures, transaction taxes are the least harmful; while they reduce welfare roughly by the amount of the tax (a reduction that can be offset by distributing tax proceeds back to traders), they do not significantly worsen market quality. Importantly, the common practice by exchanges of granting rebates to limit orders is detrimental to market quality and investor welfare, causing both higher effective spreads and longer execution times.

The paper is organized as follows. Section 3.2 provides a short overview of the existing literature. Section 3.3 describes the model and the solution methodology. Section 3.4 details our findings on the impact of HFT on market quality and investor welfare, while Section 3.5 considers the effect of proposed regulations. Section 3.6 concludes. Technical aspects of the solution methodology are described in the Appendix.

3.2. Literature Review

High-frequency trading has been a very active area of research in recent years, with the literature focusing on three main questions: the impact of HFT on market quality and welfare, the nature of HFTs' strategies, and the impact of potential regulations on markets.

Theoretical papers examining the welfare implications of HFT and its impact on other market participants have not reached a consensus on whether the effect of HFT is predominantly positive or negative. Jovanovic and Menkveld (2011) find that the presence of HFTs has an ambiguous effect on welfare. Maollemi and Sağlam (2013) show that the presence of HFTs imposes significant costs on market participants with higher latency. Along similar lines, Hoffmann (2014) finds that slow traders are left with a smaller share of overall trading surplus because of their limited ability to avoid their orders being picked-off. On the other hand, Aït-Sahalia and Sağlam (2014) show that competition among HFTs improves the welfare of slow traders and that volatility leads HFTs to reduce their provision of liquidity. Biais, Foucault, and Moinas (2015) establish that investment into fast technology may go beyond its socially optimal level.

Theoretical results on the impact of higher trading speed on market quality are also inconclusive. Using a model with endogenous information acquisition, Baldauf and Mollner (2015) show that an increase in trading speed crowds out information acquisition and causes both a decrease in the bid-ask spread and a deterioration in price efficiency. Menkveld and Zoican (2015) find that the effect of a reduction in a stock exchange's latency on the bid-ask spread is ambiguous; whether market liquidity improves or deteriorates is driven by the ratio of news traders to liquidity traders.

Several empirical papers document that HFTs improve market quality by providing the best bid-ask spreads and contribute significantly to liquidity and price discovery, and that HFT activity is not positively correlated with price and quote volatility (Brogaard, 2010, Hasbrouck and Saar, 2013, Riordan and Storkenmaier, 2012, Conrad, Wahal, and Xiang, 2015, and Hasbrouck, 2015). Hendershott, Jones, and Menkveld (2011) report that HFTs improve liquidity and enhance the informativeness of quotes, but their presence also decreases the depth of the order book and increases the costs of executing large orders. Taking execution costs into consideration, Tong (2015) measures the execution shortfall of institutional investors and finds that HFTs increase transaction costs for them. While it has been found that HFTs did not cause the Flash crash, their presence exacerbated the price movement as they absorbed immediacy ahead of others by withdrawing their orders in the face of higher market volatility (Kirilenko, Kyle, Samadi, and Tuzun, 2014).

Another strand of the literature seeks to better understand the nature of HFTs' strategies. Hagströmer and Nordén (2013) quantify the market-making activity of HFTs on the NASDAQ-OMX Stockholm exchange. They find that HFTs engaged in market making are present about 60% of the time, while other categories of HFTs are on average present only about 5% of the time. Carrion (2013) finds that in the aggregate, HFTs on the NASDAQ make money on average when supplying liquidity and lose money on average when demanding liquidity.[4] Foucault, Hombert, and Roşu (2016) investigate the impact of HFTs engaging in news-based trading on market dynamics. They find that when a HFT has faster access to news, she significantly changes her strategy and her trades account for a much larger fraction of trading volume and forecast short-run price changes. Hasbrouck and Saar (2013)

[4] Interestingly, this appears to be exactly the opposite of the findings in Brogaard, Hendershott and Riordan (2014) using the same dataset.

document that HFT leads to a high number of quote cancellations in the millisecond environment. Using a trading game with continuous prices, Baruch and Glosten (2013) find that such a strategy of fleeting orders is supported by equilibrium.

Theoretical predictions regarding the impact of current regulatory initiatives on continuous limit-order markets are scarce.[5] Harris (2013) argues that by causing liquidity-supplying HFTs to lose more often when offering liquidity, minimum resting time rules would ultimately increase investor transaction costs. Aït-Sahalia and Sağlam (2014) find that both minimum resting time rules and cancellation fees induce HFTs to quote more on both sides of the market. However, both measures lead HFTs to provide liquidity countercyclically with volatility, providing high (low) liquidity in low (high) volatility periods. Transaction taxes cause a reduction in HFT quoting and lead to lower transaction volume.

Empirical evidence on the impact of cancellation fees is provided by Malinova, Park, and Riordan (2013), who find that following their introduction on the Toronto Stock Exchange, quoting activity decreased by 30% and spreads increased by 9%. Umlauf (1993) investigates the impact of the introduction of the Swedish transaction tax and documents a large shift in trading volume from Sweden to the U.K. Colliard and Hoffmann (2015) find a large decrease in trading volume following the introduction of the French financial transaction tax in August 2012. Interestingly, the decrease is concentrated in the OTC market. While the main market remained resilient and spreads were not affected, the introduction of the tax led to a decrease in market depth, resiliency, and price efficiency.[6] Malinova and Park (2015) use a change in Toronto Stock Exchange fees to investigate the impact of make-take fees on market quality. They find that rebates improve quoted spreads, decrease adverse selection costs, and increase the aggressiveness of retail traders. They also observe that the effective spread plus total fee retained by the exchange remained constant, confirming the conjecture of Colliard and Foucault (2012) that only total fees (spread plus exchange fees) affect liquidity and trading volume. This last finding contrasts with the predictions of the model of O'Donough (2015), where the total trading cost

[5] An alternative to regulating continuous limit-order markets is to make changes to the trading mechanism as such. A few recent papers consider this approach. Budish, Cramton and Shim (2015) show that changing the design away from continuous limit order books (CLOB) to frequent batch auctions eliminates mechanical arbitrage rents that are built into the CLOB market design, enhance liquidity for investors, and stops the HFT arms race for speed. Du and Zhu (2014) characterize the socially optimal frequency of such batch auctions in a setting with private information. They show that the optimal frequency is low for scheduled information arrivals and high for stochastic information arrivals.

[6] Further results on the effect of the French financial transaction tax on market quality can be found in a number of studies. Becchetti, Ferrari, and Trenta (2014) find a reduction in turnover and intraday volatility, but mixed effects on liquidity. Meyer, Wagener, and Weinhardt (2015) document a reduction in the number of quote and price updates by liquidity suppliers and a decline in top order book depth. Gomber, Haferkorn, and Zimmermann (2015) report a drop in both liquidity demand and supply, an increase in spreads, and a decline in top order book depth.

to investors increases when the taker fee and maker rebate increase, even if the net fee is held fixed. Empirically, O'Donough reports decreasing bid-ask spreads, lower trader participation, higher order aggressiveness and higher probability of execution of limit orders as the taker fee and maker rebate increase.

Most closely related to our analysis is the recent paper by Bernales (2014). He extends the model of Goettler et al. (2009) in a similar way to ours and investigates the impact of HFT and a number of regulatory proposals on market quality. However, our analysis differs from his in several important respects. First, we model HFTs as fast speculators, i.e. as traders that have zero private valuations for the asset. By contrast, Bernales assumes that HFTs have the same distribution of private valuations as slow traders, which is unlikely to be the case in practice. Second, the set of regulatory proposals that we consider is much broader. Bernales only considers cancellation fees and an approximation for latency restrictions. Specifically, regarding the latter, in order to be able to use the original Goettler et al. (2009) model with few modifications, he models latency restrictions using a decrease in the frequency with which traders return to the market. Doing so, however, makes fast traders slower without actually enforcing these restrictions. By contrast, we model latency restrictions exactly in the way that they have been proposed for real-world markets, i.e. by either enforcing minimum resting time rules – a fixed time during which a limit order cannot be modified – or by introducing a speed bump – a delay between the time that a market order is submitted and the time it hits the order book. Modeling latency restrictions in a realistic fashion is important to be able to accurately assess their impact. Indeed, as we show below, these two forms of latency restrictions lead to different outcomes in terms of market quality and welfare. Furthermore, our analysis also considers the impact of rebate fee structures.

3.3. Model and Solution Methodology

3.3.1. Model Assumptions

Overview

We consider a continuous-time dynamic trading game in one financial asset similar to that in Goettler et al. (2009). Trading takes place by continuous double auction implemented using a limit order book. The asset's fundamental value v_t follows a random walk where the timing of price changes is driven by a Poisson process with intensity λ_v; conditional on a price change occurring, up and down moves are equally likely.

Traders arrive to the market randomly according to a Poisson process with intensity λ. They observe both the current state of the order book and the market's transaction history. We investigate both symmetric and asymmetric information settings. Under symmetric information, all traders observe the fundamental value in real time. Under asymmetric information, informed traders observe the fundamental value in real time, while uninformed traders observe it with a lag of Δ units of time and form an expectation about its current value based on their last observed value and the current state of the market.

Upon arrival, each trader selects the best action given the current state of the book, the asset's (observed or estimated) fundamental value, the transaction history and his individual parameters (described in detail below). This action may be to submit a buy or sell market or limit order or not to submit any order.

Traders whose orders are executed leave the market and never return. The others return to the market at random times which are exponentially distributed with intensity λ_r. Upon returning to the market, traders again select the best action, but based on the new state.

Structure of the Limit Order Book

We keep the notation consistent with Goettler et al. (2009). Prices are discrete $\mathscr{P} = \{p^i\}_{i=-\infty}^{+\infty}$ and the distance between any two adjacent prices is identical and called the tick size. With each price $p^i \in \mathscr{P}$ is associated a queue of outstanding limit orders denoted by l_t^i. This quantity is displayed as a positive (negative) integer for buy (sell) limit orders. The limit order book is the collection of the indexed queues $L_t = \{l_t^i\}_{i=-\infty}^{+\infty}$. The best sell order in the book is called the ask price and denoted $A(L) = p^{\min\{i|l_t^i<0\}}$. Analogously, the best buy order is called the bid price and given by $B(L) = p^{\max\{i|l_t^i>0\}}$. In the implementation, we use a grid with a finite number of prices n_p and center these around the fundamental value v_t. We choose the number of ticks n_p sufficiently high that orders never "fall off" the grid. That is, orders are revised by returning traders before becoming too unaggressive for the grid, or get picked off before becoming too aggressive for the grid.

As is standard in limit order markets, the priority of execution of limit orders is based on price and time. Price priority means that buy (sell) orders at higher (lower) price have priority over those at lower (higher) prices. Time priority gives the limit orders in the particular price queue preference depending on the time of arrival. When an order executes against an outstanding limit order, the order which was submitted earlier at that price has priority. A buy (sell) market order is an order which executes against the resting sell (buy) limit orders starting from the best ask (bid). A marketable limit order is the same as a market order with a specified maximum (minimum) price for buying (selling) a specified number of units of the asset.

Trader Types and Behavior

New traders arrive to the market according to a Poisson process with intensity λ. Upon arrival, they may choose to submit a buy or sell market or limit order or not to submit any order. The action they select is optimal given the current state of the limit order book, the history of transactions, the asset's (observed or estimated) fundamental value, and their type. Changes in the fundamental value expose limit orders to two risks: (i) obtaining an undesirable execution when the fundamental value moves in an unfavourable direction, and (ii) not obtaining a desirable execution when the fundamental value moves in a favourable direction.

Traders may submit orders only for one share. Traders whose order is executed leave the market. The others return to the market with intensity λ_r. The reentry (or monitoring) rate λ_r depends on the technology available to the trader – it will be high when monitoring costs are low as in Foucault, Kadan, and Kandel (2013) – and represents a friction which traders have to take into account when making their decisions. A higher monitoring rate allows faster reaction to changing market conditions, giving the trader the ability to execute against a stale limit order or to cancel or resubmit her own order and thus avoid her order being picked-off.

Upon reentry, traders choose their action optimally, depending on the current state of the limit order book, the history of transactions, the asset's (observed or estimated) fundamental value, their type, and the priority and position of their last limit order if they submitted any. They are allowed to retain their existing limit order and they do so if this is the highest expected value action for them. Otherwise, they might cancel the existing order, possibly incurring a cancellation cost and choose a different value-maximizing action.

Traders' type θ is defined by three attributes, $\theta = \{h, \alpha, \rho\}$: whether they are HFT or not ($h \in \{0, 1\}$), their private valuation for the asset, α, and their impatience, ρ. The private valuation α is drawn from a discrete probability distribution in which the probability of a zero private value is positive. Traders with a zero private value are called speculators, those with a nonzero private value are investors. Traders with nonzero private valuation α seek to materialize this potential benefit through trade.

Traders' type is also defined by their impatience parameter ρ, which enters their utility function in a similar fashion as a discount factor. The impatience coefficient represents the trader's disutility of obtaining execution later. Traders are risk neutral and their instantaneous utility at time t is given by:

$$u_t = \begin{cases} \alpha + v_t - p^i & \text{if she executes a buy order at price } p^i \text{ and time } t, \\ p^i - \alpha - v_t & \text{if she executes a sell order at price } p^i \text{ and time } t, \\ 0 & \text{if she does not execute at time } t. \end{cases} \tag{3.1}$$

HFTs are modeled as fast speculators, i.e. their monitoring rate exceeds that of other traders, $\lambda_r^h > \lambda_r$, and their private valuation for the asset is zero.[7] In addition, when investigating asymmetric information settings, we assume that HFTs observe the fundamental value v_t in real time, whereas slow traders observe it with a lag of Δ time units as in Foucault, Hombert, and Roşu (2013). Thus, in the presence of asymmetric information, non-HFTs have to form an expectation about the current fundamental value v_t based on the lagged fundamental value $v_{t-\Delta}$ and the current state of the market in the same way as the uninformed traders in Goettler et al. (2009).

3.3.2. Time Line and Solution of the Game

Since the game evolves in continuous time and the different events that can occur arrive stochastically, there is no pre-set sequence of events. Rather, when solving the game, the first step is to draw the time of the next event from an exponential distribution with intensity given by the overall event arrival intensity Λ_t, which is the sum of (i) the intensity of arrivals of new traders λ, (ii) the weighted intensity of arrivals of returning traders λ_r and λ_r^h, and (iii) the intensity of changes in the fundamental value λ_v. Formally,

$$\Lambda_t = \lambda + N_t^{h=0}\lambda_r + N_t^{h=1}\lambda_r^h + \lambda_v, \tag{3.2}$$

[7] A categorization of traders based on their intrinsic motivation to trade, presence in the market, information, trading strategies and other characteristics is provided in Harris (2002). Using this categorization, we distinguish two basic categories of traders – those with an intrinsic motivation (investors) and those without (speculators). We model HFTs as fast speculators who may have an information advantage because of faster information processing.

where $N_t^{h=0}$ denotes the number of slow traders and $N_t^{h=1}$ the number of HFTs that have not left the market at time t.

Once the time of the next event is known, its nature – new arrival, return, or change in the fundamental value – is determined randomly based on the ratio of the intensity of that event to the overall event intensity Λ_t. The game is then played following the workflow summarized in Figure 3.1. Specifically, when a new trader arrives, the trader performs her optimization and submits her order or no order. When an existing trader returns, the trader performs her optimization and retains or cancels her existing order and, in the latter case, submits a new order or no order. All active orders are stored in the limit order book and if a new order was submitted, it enters in the precise position specified by the trader. If the trader submits a market order, the order executes against a resting limit order; that order is then removed from the book and the priorities of the remaining limit orders are updated. Traders whose orders are executed are removed from the system. Finally, if the event is a change in the fundamental value v_t, the price grid in the limit order book is shifted as described in Section 3.3.1. Throughout, payoffs, transaction costs and other statistics are computed and stored.

Since traders have privately known α's as well as possibly private information about the fundamental value v_t and time is not a state variable, the solution concept for this game is Markov-perfect Bayesian equilibrium as pointed out in Maskin and Tirole (2001). We focus on symmetric equilibria, where traders of the same type optimally choose the same action in a given state.

The state space consists of the past history of the game, the current limit order book, the trader's type and the position of his past order (if any) in the book. Our state space is richer than that in Goettler et al. (2009) because of the additional parameter h. Let Θ denote the set of feasible agent types. Ideally, the state space considered during decision making would comprise all the variables of the limit order book and the entire transaction history. In order to make the problem tractable, we make certain restrictions. Specifically, we do not consider the entire transaction history but just the last transaction, whether it was buyer- or seller-initiated, and the transaction price. Moreover, we do not consider the full detail of the limit order book, but focus on its most informative variables, namely bid price, ask price, bid size, ask size, depth off bid and depth off ask. We perform tests to determine whether there is a significant difference in actions compared to those arising when using a broader information set, but like Goettler et al. (2009) find that broadening the state space does not improve predictive power. A mathematical description of the state space and the optimization problem faced by traders is provided in the Appendix.

3.3.3. Parameterization

In setting parameter values, we relate to Goettler et al. (2009) and the empirical literature cited therein in order to ensure the comparability of our results. We set the arrival rate for new traders λ to 0.25. The returning rate of HFTs λ_r^h is set to 4 and that of non-HFTs λ_r to 1. Thus, we assume that HFTs react four times faster than non-HFTs.

The discount rate ρ is set to 0.03 for all traders. The support of the private valuation α in ticks is $\{-4, -2, 0, 2, 4\}$ with probability distribution $F_\alpha = \{0.15, 0.35, 0.65, 0.85, 1.0\}$. This private value distribution roughly corresponds to the empirical findings in Hollifield, Miller, Såndas and Slive (2006) for the Vancouver stock exchange.

The tick size is set to 1/8. The expected time between changes in the fundamental value is set to 10 units of time, i.e. $\lambda_v = 0.1$. When the fundamental value changes, it always does by one tick, with up and down moves equally likely. When considering asymmetric information settings, we set the lag Δ with which slow traders observe the fundamental value to 2 units of time; we experimented with higher values and found that the results do not change significantly.

The limit orders being tracked in the book may lie between 7 ticks below and 7 ticks above the fundamental value. The fundamental value lies on the tick in the middle of this price range. Limit orders may be submitted at prices between 3 ticks below and 3 ticks above the fundamental value. There are thus seven possible prices at which a limit order can be submitted.

3.4. The Impact of HFT on Market Quality and Welfare

In this section, we investigate the impact of HFT on market quality and welfare by comparing the properties of equilibrium between a situation without HFTs and situations with HFTs. We assess market quality using the following standard measures:

1. The *bid-ask spread* is defined as the difference between the best ask price and the best bid price. It measures the price for immediacy.

2. The *effective spread* is defined as the absolute value of the difference between the transaction price and the mid-quote at the time of the transaction, and measures effective trading costs.

3. *Market depth* is measured by the number of units of the asset available to buy or sell in the limit order book. We consider both the depth at bid and ask and the depth off-bid and off-ask. Large depth means that the price impact of an order of a given size and thus related transaction costs are small.

4. *Price discovery* measures the percentage distance between the fundamental value and transaction prices. In efficient and transparent markets, the difference between the fundamental value and transaction prices is small.

5. *Microstructure noise volatility* measures the dispersion of transactions around the asset's fundamental value. The higher microstructure noise volatility, the larger execution risk for traders, for example because their limit orders may be picked off or become stale.

To assess the impact of HFT on welfare, we compute each trader's utility using equation (Equation (3.1)), i.e. as the difference between his valuation for the asset (fundamental value plus private valuation) and the price at which he is able to purchase or sell the asset on the market. For limit orders, utility is discounted by the amount of time it took to execute the limit order. We then compute welfare as the average utility of all traders in the game. Since regulators are likely to be more concerned about the welfare of investors than that of speculators, in addition to total welfare, we also compute welfare separately for investors and speculators.

Before we proceed, it is useful to consider the trade-offs faced by traders in the market and how these might be affected by the presence of HFTs. We say that an order is more aggressive if it is submitted at a price which is closer to the opposite side of the market. The most aggressive order is a market order. The more aggressive an order is, the higher the execution probability prior to the trader returning. This probability, however, also depends on the other orders already in the book and on which new orders are submitted before the trader returns. As mentioned earlier, changes in the fundamental value expose limit orders to two risks: (i) obtaining an undesirable execution when the fundamental value moves in an unfavourable direction, and (ii) not obtaining a desirable execution when the fundamental value moves in a favourable direction. In the first case, the presence of HFTs exacerbates picking-off risk for a slow trader because HFTs are likely to react to the new situation before the slow trader's returning to the market and lift the order. In the second case, the presence of HFTs reduces the probability that a slow trader wishing to update her order given the new situation will be able to pick off a too aggressively priced order on the other side of the market. Furthermore, if she decides to submit a limit rather than a market order when updating her order, that order is likely to lie behind HFTs' orders in the queue, delaying execution.

When conducting our quantitative analysis, we model the situation without HFTs using the base case parameter values described in Section 3.3.3. Recall that in this parameterization, the shares of traders with private valuations of 0, 2, and 4 ticks in absolute value are 30%, 40%, and 30%, respectively. We assess the impact of HFT in two ways: (1) by gradually replacing slow speculators with HFTs, and (2) starting from a situation in which half of speculators are slow and the other half are HFTs (i.e. a situation in which both groups each make up 15% of the overall population), by gradually adding HFTs until their overall share of the population reaches 25%. Note that while the first approach only affects the distribution of traders, the second increases overall trader arrivals and therefore overall market activity. Arguments could be made in favor of using one or the other approach; as will be shown shortly, however, our findings are robust to the approach chosen.

The impact of HFT on market quality and welfare using the first approach can be seen in Table 3.1 and Figure 3.2 for the symmetric information case and Table 3.3 and Figure 3.6 for the asymmetric information case. The results using the second approach are reported in Table 3.2 and Figure 3.4 for the symmetric information case and Table 3.4 and Figure 3.7 for the asymmetric information case. We highlight our main observations regarding market quality and investor welfare and then discuss them in more detail. Throughout, we begin with the results under symmetric information and then contrast them with those under asymmetric information.

Observation 1 (Market Quality)*: The presence of HFTs increases market depth, reduces spreads and transaction costs, and improves price discovery. Under asymmetric information, these effects arise only if the share of HFTs in the market is sufficiently large to generate competition between them.*

Consider first the case where slow speculators are replaced with HFTs. As can be seen in the top panel of Figure 3.2, an increase in the share of HFTs among speculators significantly improves depth and yields small improvements in the quoted bid-ask spread, the effective spread, and price discovery. The volatility of microstructure noise is also slightly lower (see Table 3.1). Two channels drive these results. First, HFTs' shorter reaction times allow them to update their limit orders faster, resulting in quotes that track the evolution of the fundamental value more closely and improving price discovery. HFTs' faster reaction times also leads to lower picking-off risk. Accordingly, as is apparent in the bottom panel of Figure 3.2, the expected payoff of their limit orders is significantly higher than for slow speculators. Second, HFTs are faster at picking-off mispriced orders using market orders. This results in their submitting a larger proportion of market orders than slow speculators, as can be seen in the bottom panel of Figure 3.2. Although profits from lifting mispriced orders remain constant at one tick (0.125, see the bottom panel of Figure 3.2, right axis), an increase in the proportion of HFTs increases competition, reducing the probability of lifting mispriced orders. This induces both HFTs and slow speculators to reduce their use of market orders and leads to higher liquidity provision by both groups.[8] With increasing competition, however, pick-off risk increases for both groups of speculators, as is apparent from the decrease in their utility from limit orders in the bottom panel of Figure 3.2. A closer look at the bottom panel reveals that slow speculators' expected payoff from limit orders in the case without HFTs (the intercept of the solid blue line) exceeds HFTs' expected payoff from limit orders in the case where there are no slow speculators (the value of the dashed blue line on the right). This means that investors demanding liquidity pay a lower price in the case where all speculators are HFTs than in the case without HFTs.

Inspection of traders' strategies in Table 3.1 reveals that consistent with the findings in Hoffmann (2014), tighter spreads cause slow traders to submit less aggressive orders as competing for the best quotes would increase their risk of being picked-off. Tighter spreads also reduce the volatility of microstructure noise. As their share in the market increases, HFTs provide more liquidity by increasing their share of limit orders (see the bottom panel of Figure 3.2), which benefits investors in the form of shorter execution times (see the *Transactions* section in Table 3.1). HFTs' ability to submit high priority orders also results in an increase in depth at the bid and ask and further in the book. These findings are in line with the empirical studies by Hasbrouck and Saar (2013) for U.S. data and Riordan and Storkenmaier (2013) for German data. As can be seen in Figure 3.3, in the presence of HFTs, the order book is more skewed towards the fundamental value (located in the middle of the price grid).

[8] Bernales (2014) also finds that HFTs favor market orders when their share of the market is small and limit orders when it is large. However, he does not provide a breakdown of the order types for the different groups of slow traders and only considers an asymmetric information setting.

As can be seen in Table 3.2 and Figure 3.4 and Figure 3.5, under symmetric information, the impact of HFT on market quality is similar if HFTs are added to the market instead of substituting slow speculators.

The impact of HFT on market quality is less clear-cut under asymmetric information. This is best seen in the case where the overall trader population is constant (see Table 3.3 and Figure 3.6). Since slow traders observe the fundamental value with a lag, they must base their trading decisions on an estimate of the fundamental value rather than on the true value. This gives HFTs better opportunities to pick-off mispriced orders. Imperfect information about the fundamental value also causes slow traders to make mistakes when submitting market orders to lift orders they believe to be mispriced, resulting in a growing difference between their and HFTs' payoff from market orders (see the bottom panel of Figure 3.6). The beneficiaries of these mistakes are HFTs, whose average discounted payoff from limit orders increases compared to the symmetric information case (contrast the bottom panels of Figure 3.2 and Figure 3.6; the exact values are 0.0348 under symmetric information and 0.0448 under asymmetric information). When the share of HFTs in the market is small, the adverse selection causes price discovery and spreads to worsen. As the share of HFTs in the market increases, however, competition among them leads to an improvement in all market quality measures. When the share of HFTs is sufficiently large, the competition effect dominates the adverse selection effect and market quality is better than in the case without HFTs. In the case where HFTs are added to the market instead of substituting slow speculators (see Table 3.4 and Figure 3.7), the share of HFTs is already sizable at 15% initially. As a result, the competition effect dominates and market quality improves as HFTs are added.

It is instructive to contrast our results with those in Bernales (2014). He investigates the asymmetric information case and finds that market quality increases monotonically with the share of HFTs in the market. Our results show that while this statement holds under symmetric information, it does not hold under asymmetric information. Rather, when HFTs have an informational advantage, HFT initially reduces market quality, and increases it once HFTs' share in the market becomes relatively large.[9]

[9] The reason that Bernales (2014) identifies a monotonic impact is that he considers step sizes for the market share of HFTs of 20% and the nonmonotonicity occurs in the range between 0% and 20%.

***Observation 2 (Welfare)**: HFT does not reduce investor welfare but affects the welfare of slow speculators negatively. HFTs' welfare is significantly higher than that of slow speculators. HFTs supply more liquidity than investors, but less than slow speculators.*

Turning to investor welfare and again starting with the case where information is symmetric and slow speculators are replaced with HFTs, we find that both overall welfare and average welfare for all slow traders (speculators and investors) are not negatively affected by the presence of HFTs (see Figure 3.2, middle panel). However, the impact of HFTs on welfare differs across trader categories. Whereas slow speculators' welfare falls markedly, investors experience a slight increase in welfare reflecting the benefits from increased liquidity in the market. Because of improved liquidity and the increase in picking-off risk associated with HFT, investors' use of market orders rises (see the *Trader Strategies* section in Table 3.1).

Reflecting their speed advantage, HFTs earn higher surpluses than slow speculators. Both aspects of slow speculators' disadvantage compared to HFTs are visible in Table 3.1: since they are slow at updating their orders when the fundamental value changes, (i) they earn lower profits on limit orders than HFTs, and (ii) they seldom manage to pick off mispriced orders, as can be seen from the lower share of market orders in their transactions compared to HFTs.

Interestingly, the results in Table 3.1 also reveal that HFTs supply more liquidity than investors, but less than slow speculators. As the share of HFTs increases, slow speculators initially substitute market orders with aggressive limit orders. As the share of HFTs increases further, however, competing for the best quotes with HFTs in an environment with narrower spreads increases slow speculators' pick-off risk, inducing them to submit fewer aggressive limit orders. Slow speculators provide liquidity even more often than HFTs, but their technological disadvantage prevents them from earning substantial payoffs as their orders are frequently picked-off. As one would expect, as the share of HFTs among speculators rises, increased competition between them erodes their trading profits. The decline in profits, however, is even more pronounced for slow speculators.

As can be seen in Table 3.2 and Figure 3.4 and Figure 3.5, the impact of HFT on welfare is similar if HFTs are added to the market instead of substituting slow speculators. Thus, the impact of HFT on both market quality and welfare is robust to the way that it is modeled.

Under asymmetric information (see Table 3.3 and Figure 3.6), the impact of HFT on welfare is similar to that under symmetric information: the welfare of investors increases slightly, and that of slow speculators falls. It is worth noting that slow speculators' welfare falls more strongly under asymmetric information than under symmetric information because they face both a speed and an informational disadvantage. Similar patterns can be observed if HFTs are added to the market instead of substituting slow speculators (see Table 3.4 and Figure 3.7).

Thus, our results highlight that the welfare impact of HFT on slow traders is heterogeneous. While slow speculators generally experience a welfare reduction from increased competition by HFTs, the welfare of investors increases. Put differently, the negative impact of HFT on slow trader welfare identified by Bernales (2014) does not hold for all subgroups, but only for slow speculators.

3.5. Impact of Proposed Regulations

In this section, we investigate the effectiveness of a number of regulatory measures that have been suggested to improve market quality in the presence of HFT. We focus on the main regulatory initiatives, namely minimum resting time rules, cancellation fees, transaction taxes, make/take (rebate) fee structures and speed bumps. We compute the equilibria arising when introducing each of these regulations, thus fully accounting for the resulting changes in traders' optimal strategies, and collect the market quality and welfare statistics. We then assess the impact of each regulation by comparing the properties of the resulting equilibrium with those of the unregulated case. Section 3.5.1 describes the different regulations and how they are implemented in our analysis, and Section 3.5.2 presents our findings.

3.5.1. Description of the Regulations and Implementation in the Model

Minimum resting time

Minimum resting time (also known as minimum holding or waiting time and henceforth abbreviated as MRT) is an artificial delay imposed between the time an order is submitted and the time at which it may be changed or cancelled. According to D'Antona (2010), the topic of a minimum resting time for quotes came up during joint meetings on the Flash crash between the SEC and the Commodity Futures Trading Commission that took place in summer 2010. The range of values that were considered in these discussions was from 50 milliseconds to one second. Such a measure would in effect impose on all market participants some of the obligations that are imposed on market makers by some exchanges.[10] The inability to immediately change orders increases picking-off risk for traders, thus making it more costly to submit limit orders. The implementation of MRT rules in the model is straightforward. A trader that had submitted an order at time s and returns at time $t < s + MRT$, i.e. before the minimum resting time requirement is met, is put into a waiting queue. Further events in the game are then drawn and processed as long as their time is less than $s + MRT$. As soon as an event time after $s + MRT$ is drawn, the trader recorded in the queue is allowed to modify or cancel her order based on the state at time $s + MRT$. We report results for a MRT of one time unit. We have investigated MRT values between 0.25 and five time units and came to similar conclusions.

Cancellation fee

A cancellation fee is a fee that traders must pay whenever they cancel or modify their orders. Cancellation fees make updating orders costly and must be taken into consideration by traders when choosing the aggressiveness of their limit orders. In the implementation, the fee is subtracted from the trader's current utility whenever an order is cancelled or updated. Note that any fees incurred following potential later order cancellations are automatically included in the continuation utility associated with the current state by the solution algorithm and therefore do not need to be tracked explicitly. To make the welfare comparison with the unregulated situation fair, fee proceeds are then distributed equally across all traders. For completeness, we report total welfare both without and with fee proceeds. We report results for a cancellation fee value of 0.015, corresponding to roughly 0.38% of the average asset price.[11] We also experimented with values between 0.001 and 0.5 and found the results to be similar.

Transaction tax

Under a transaction tax, traders have to pay a tax amount equal to a certain proportion of the value of their transactions. In the implementation, the tax is collected at the time of order execution and, consistent with taxes typically imposed in the real world, shared equally between the seller and the buyer. To make the welfare comparison with the unregulated situation fair, tax proceeds are then

[10] For example, in addition to submitting quotes on both sides of the market, market makers on the Eurex options exchange have to quote with a minimum holding time of ten seconds. Similar rules are in force on the Hong Kong Exchange and on the ASX options market, where the holding period is 30 seconds.

[11] According to Hollifield et al. (2006) and Goettler et al. (2009), based on the distribution of private values that we use, the expected stock price is roughly 3.89. Because private valuations are additive in the utility function and orders do not "fall off" the price grid, traders only consider changes in the fundamental value, not its level, which follows a random walk and has an infinite support.

distributed equally across all traders. For completeness, we report total welfare both without and with tax proceeds. We set the tax rate at 0.1%, a value in the ballpark of rates imposed in several countries. Because the fundamental value follows a random walk and transactions happen at prices around the fundamental value, the transaction tax lowers agents' final payoff by 0.001945. We also investigate varying the tax rate from 0.05% to 0.5% and come to similar conclusions.

Make/take fees

Under a make/take (or rebate) fee structure, whenever a transaction is executed, the consumer of the liquidity has to pay a taker fee and the trader who provided the limit order receives a maker fee (a rebate on transaction fees). As mentioned in the Introduction, such fee structures are used by many exchanges in order to encourage liquidity provision. However, their desirability is controversial.[12] Since such fee structures discourage trading by market orders and encourage submitting limit orders, one would expect them to result in tighter quoted spreads. To implement rebates in the model, we simply subtract the taker fee from the utility of the trader that consumed liquidity and add the maker fee to the utility of the trader whose limit order got picked off. We set the taker fee at $0.01 and the maker fee at (minus) $0.01, which is in line with the actual pricing structures at various exchanges. We set both fees symmetrically in order to make the fee structure budget neutral overall, thus facilitating the comparison with the unregulated case. We investigated the properties of market equilibrium for a range of make/take fee structures, both symmetric and asymmetric, and obtained similar results.

Speed bump

A speed bump is an artificial delay imposed on incoming orders; when submitted, orders are stored for a fixed amount of time before they are routed to the order book. Traders submitting orders have to take into account that by the time their order reaches the book, the book could be different than what they observed when they submitted the order. As mentioned in the Introduction, speed bumps have been imposed on all orders on some exchanges and only on market orders on others. In our analysis, we choose to apply the speed bump to market orders only. Such orders are converted to marketable limit orders upon submission in order to protect traders against adverse quote movements between the time that an order is submitted and the time it hits the order book. However, execution is no longer

[12] For example, in its Concept Release on Equity Market Structure, the SEC (2010) raises the following concerns: "Are liquidity rebates unfair to long-term investors because they necessarily will be paid primarily to proprietary firms engaging in passive market making strategies? Or do they generally benefit long-term investors by promoting narrower spreads and more immediately accessible liquidity? Do liquidity rebates reward proprietary firms for any particular types of trading that do not benefit long-term investors or market quality?"

guaranteed. If the best quote the market order intends to trade on moves in an unfavourable direction, the order is recorded in the intended position in the book and the trader can decide on her optimal action again upon her next reentry. This is consistent with the way that speed bumps have been implemented on Nasdaq. We report results for a speed bump of one time unit. We have experimented with values between 0.25 and five time units and found similar results.

3.5.2. Results

We conduct our analysis using the parameterization described in Section 3.3.3. Recall that in this parameterization, the shares of traders with private valuations of 0, 2, and 4 ticks in absolute value are 30%, 40%, and 30%, respectively. We assume that half of speculators are HFTs. As mentioned at the beginning of this section, we assess the impact of each regulation by comparing the properties of the resulting equilibrium with those of the unregulated case. We consider both the symmetric information case, where all traders observe the fundamental value in real time, and a situation with asymmetric information where HFTs observe the fundamental value in real time and all other traders observe it with a lag of two time units. The market quality and welfare indicators under symmetric and asymmetric information are reported in Table 3.5 and Table 3.6, respectively. We highlight our main observations regarding market quality and investor welfare and then discuss them in more detail.

Observation 3 (Minimum Resting Time)*: MRT does not improve market quality, reduces aggregate welfare, transfers welfare from HFTs to slow speculators, increases the aggressiveness of orders submitted by investors and increases order execution time for all trader groups.*

Overall, MRT has a mixed impact on market quality. It reduces the depth of the book, but leads to a decrease in both quoted and effective spreads in the presence of asymmetric information.[13] At the same time, MRT worsens price discovery and increases the volatility of microstructure noise under both symmetric and asymmetric information. This finding is in line with the results in Kirilenko and Lamacie (2015), who report that an increase in latency leads to higher short-term volatility.

MRT has a negative impact on aggregate welfare. Importantly, the decrease in aggregate welfare is entirely borne by investors and not by speculators. The welfare of HFTs does decrease, but this

[13]While our results on market depth are in line with those reported by Bernales (2014) who analyzes the asymmetric information case, our results on spreads differ from his. As mentioned in the introduction, while we implement MRT rules exactly in the way that they have been proposed for real-world markets, Bernales (2014) models them by reducing the returning frequency of fast traders. Doing so makes fast traders slower without actually enforcing the MRT restriction.

decrease is exactly offset by an increase in the welfare of slow speculators, which are the only group that is better of under MRT rules. The reason for the welfare transfer from HFTs to slow speculators is that HFTs' inability to modify or cancel their outstanding limit orders sufficiently fast exposes them to increased picking-off risk. HFTs' reduced reaction ability is clearly visible in Table 3.5 and Table 3.6: the share of limit orders that they cancel falls by an order of magnitude, and the fraction of their transactions from limit orders rises (equivalently, the fraction of transactions that they initiate via market orders falls).

Due to increased picking-off risk, HFTs substitute risky aggressive limit orders with less risky orders at the best quote. By contrast, slow speculators and 4 ticks private value investors, who are now better able to pick-off stale orders, are more likely to submit market orders. Overall, HFTs end up providing more liquidity. However, the fact that liquidity provision becomes more risky for all trader groups results in wider spreads. This lower quality of liquidity reduces liquidity consumption. Accordingly, as can be seen at the bottom of Table 3.5 and Table 3.6, execution times increase for all groups of traders, negatively affecting welfare.

Observation 4 (Cancellation Fees): *Cancellation fees decrease market quality and welfare, increase the usage of market orders and limit orders posted at the best quote, and lead to a rise in order execution time for all trader groups.*

Consistent with the empirical findings in Friederich and Payne (2015), cancellation fees decrease depth and worsen price discovery. Under symmetric information, the quoted spread remains the same and the effective spread worsens. Under asymmetric information, the quoted spread falls, but the effective spread remains the same. Thus, even ignoring the fee itself, the introduction of cancellation fees does not reduce transaction costs.

Turning to welfare, cancellation costs lower overall welfare; importantly, distributing fee proceeds to traders does not offset the welfare loss. Welfare losses are registered in all trader groups, but slow speculators facing asymmetric information are least affected. As one would expect, order cancellation ratios fall. More strikingly, cancellation fees lead to a sharp decrease in spread-improving (i.e. aggressive) limit orders: realizing that the fees make it more costly to remove orders that have become mispriced to prevent them from being lifted, traders become more prudent in their order submission. Instead of aggressive limit orders, they use market orders or limit orders submitted at the best current quote. Another consequence of cancellation fees is that they induce traders to decide not to submit any order initially and wait until they return to do so, leading to lower overall depth in the book. This causes liquidity to become hidden and execution times to rise significantly for all trader groups. The reason is similar to that in the case of MRT rules: cancellation fees make liquidity provision more costly for all trader groups, leading to lower quality liquidity and reduced liquidity consumption.

Observation 5 (Transaction Taxes)*: Transaction taxes reduce the trading gains of all trader groups roughly by the amount of the tax and do not significantly impact market quality, aggregate welfare, and trading strategies.*

Transaction taxes only have a small impact on market quality. Welfare decreases for all groups of traders roughly by the amount of the tax; accordingly, distributing tax proceeds to traders completely offsets the welfare loss. The decrease in welfare is somewhat stronger for slow speculators because the tax represents a larger portion of their welfare without the tax. Under asymmetric information, transaction taxes lead uninformed traders to submit fewer aggressive limit orders because the riskiness of such orders remains the same but their payoff is lower.

It is worth noting that with a finer distribution of private values (or with a higher tax rate), some traders would decide not to trade at all, which would lead to wider spreads and a decrease in market liquidity as was observed by Colliard and Hoffmann (2015) in their event study on the introduction of the French financial transaction tax.

Observation 6 (Rebates)*: Rebates reduce market quality and investor welfare, increase the aggressiveness of speculators, and lead to a rise in order execution time for all trader groups.*

Contrary to what one might expect, rebates lead to worse price discovery and reduce depth. Although they improve the quoted spread in the symmetric information case, this does not translate into lower transaction costs as measured by the effective spread. Rather, the effective spread increases under both symmetric and asymmetric information.

Rebates also lead to a reduction in overall welfare. They increase the welfare of slow speculators at the expense of investors, while HFT welfare is almost unaffected. The higher costs of market orders lead traders to avoid such orders and induces longer execution times of limit orders for all groups of traders. Slow speculators and especially HFTs do not compete for the best quotes (the share of aggressive limit orders that they submit is lower than in the base case). Rather, they wait for investor quotes to become mispriced and lift them, as can be seen from the higher proportion of market orders in speculators' strategies and transactions. As HFTs can react to mispriced limit orders faster, the proportion of market orders in their transactions rises more strongly than for slow speculators. Like speculators, investors also submit fewer aggressive limit orders and more limit orders at the best quote than in the base case. However, the proportion of limit orders at the best quote increases more strongly for investors than for speculators, and a sizable part of that increase comes from a reduction in orders submitted below the best quote. The reason that investors increase their use of orders at the best quote is that submitting market orders is costly because of the take fees and submitting aggressive limit orders is risky due to speculators' more frequent lifting of mispriced orders. Accordingly, rebates cause market depth to fall more strongly off the best quote than at the best quote.

Observation 7 (Speed Bumps)*: Speed bumps for market orders increase microstructure noise and worsen price discovery. They reduce the welfare of slow speculators, but do not affect investor welfare. Speed bumps also make order submissions by investors and HFTs more aggressive and increase order execution time for all trader groups.*

Speed bumps generally worsen price discovery, but improve spreads and depth in the book. The welfare of investors is unaffected and that of slow speculators decreases. Speed bumps render the execution of market orders uncertain. Perhaps surprisingly, speed bumps actually lead to an increase in market order submissions in all groups. The reason is that since speed bumps for market orders are implemented as marketable limit order with the limit set at the current best limit order on the other side of the market (consistent with the way that speed bumps have been implemented on Nasdaq), traders might obtain slightly better execution. This induces them to submit blind orders and leads to higher pick-off risk for standing limit orders. This is reflected in lower cancellation ratios for HFTs and a higher proportion of limit orders in their transactions.

The results in Table 3.5 and Table 3.6 relate to a speed bump with a deterministic delay. Harris (2013) proposed introducing speed bumps with random delays, primarily to deter the technology arms race between HFTs. In order to assess whether such random speed bumps lead to improvements in market quality and welfare, we implemented two versions of his proposal. In the first, the speed bump was drawn from a uniform distribution between zero and one second, while in the second, it was drawn from a uniform distribution with strictly positive support (the cases considered were between 0.1 and 1.1 second and between 0.5 and 1.5 seconds). All these cases led to results that were similar to those for deterministic speed bumps presented in Table 3.5 and Table 3.6 and are not reported for brevity.

We ran numerous additional cases, varying both the parameter values of the underlying base case (in particular the volatility of the asset's fundamental value and the proportions of investors, slow speculators and HFTs) and the values of the parameters capturing the different regulations. The general conclusion that none of these regulations consistently improves investors' welfare is robust.

3.6. Conclusion

We investigate the impact of HFT and a number of regulatory initiatives on market quality and investor welfare using a general dynamic limit order book model as in Goettler, Parlour, and Rajan (2009). We solve for the Markov perfect equilibrium numerically using the Q-learning algorithm. We find that the presence of HFTs improves most common measures of market quality, reducing spreads, increasing market depth, and enhancing price discovery. These results hold without qualification under symmetric information. When HFTs have an informational advantage over slow traders, however, improvements in market quality arise only if competition between them is sufficiently strong. We also find that HFT does not negatively impact investor welfare. However, it leads to a reduction in the welfare of slow speculators, who are crowded out by HFTs.

The flexibility of the model allows us to investigate the impact of the main recent regulatory initiatives designed to curb HFT on market quality and investor welfare. We consider minimum resting time rules, cancellation fees, transaction taxes, rebate fee structures and speed bumps. We find that most of these regulatory proposals have a negative impact on market quality and that none consistently improves investor welfare. Admittedly, some of these regulatory proposals lead to improvements in a number of market quality measures; for example, minimum resting time rules reduce trading costs under asymmetric information, while speed bumps increase book depth and decrease spreads. However, these improvements generally do not translate into higher welfare for long-term investors. Rather, the main effect of these regulations is to generate wealth transfers from HFTs to slow speculators. These regulations therefore appear inadequate to enhance investor welfare in the presence of HFTs. Of the different measures, transaction taxes are the least harmful; while they reduce welfare roughly by the amount of the tax (a reduction that can be offset by distributing tax proceeds back to traders), they do not significantly worsen market quality. Importantly, the common practice by exchanges of granting rebates to limit orders is detrimental to market quality and investor welfare, causing both higher effective spreads and longer execution times.

It should be noted that while we have considered asymmetric information between HFTs and other traders in our analysis, we have assumed that markets are not fragmented and ruled out front-running of already submitted orders by HFTs. In order to make the analysis tractable, we have restricted individual traders to trade a single share. A promising avenue for future research is to investigate to what extent our findings generalize to settings where traders may submit large orders, thus requiring them to incorporate the issues of price impact and detection by other traders in their decision-making process.

Table 3.1

Impact of HFT on market quality, welfare, and strategies when the overall trader population is constant under symmetric information.

Share of HFTs among Speculators	0%	25%	50%	75%	100%
Market Quality					
Best Bid/Offer size	1.419	1.708	1.976	2.282	2.569
Depth off BBO	4.611	5.024	5.822	6.447	7.558
Quoted spread	1.610	1.611	1.615	1.575	1.575
Effective spread	1.266	1.298	1.176	1.100	1.079
Price discovery $\|p_t - v_t\|/v_t$ %	1.322	1.315	1.289	1.260	1.166
Microstructure noise volatility	0.080	0.080	0.080	0.079	0.077
Welfare					
Average welfare	0.235	0.237	0.239	0.240	0.242
Slow speculator	0.069	0.060	0.050	0.035	-
Investor 2 ticks PV	0.216	0.217	0.219	0.222	0.228
Investor 4 ticks PV	0.434	0.434	0.435	0.436	0.440
HFT	-	0.105	0.096	0.084	0.066
Trader Strategies					
Slow speculator					
Buy market order %	0.153	0.103	0.064	0.046	-
Aggressive buy limit order %	1.282	1.337	1.454	1.200	-
At quote buy limit order %	9.946	11.476	10.779	10.147	-
Below best buy limit order %	38.990	40.100	36.537	37.694	-
Investor 2 ticks private valuation					
Buy market order %	5.570	5.416	5.420	5.736	5.903
Aggressive buy limit order %	10.027	10.450	13.777	16.736	17.940
At quote buy limit order %	45.644	46.547	39.156	31.097	22.696
Below best buy limit order %	34.607	34.243	36.161	42.232	50.017
Investor 4 ticks private valuation					
Buy market order %	18.159	17.558	16.951	16.754	14.030
Aggressive buy limit order %	16.893	16.515	17.898	18.841	18.492
At quote buy limit order %	55.195	56.216	53.357	47.620	45.845
Below best buy limit order %	9.029	8.986	10.880	15.968	21.633
High-frequency trader					
Buy market order %	-	0.073	0.047	0.033	0.022
Aggressive buy limit order %	-	1.026	1.106	0.845	0.779
At quote buy limit order %	-	8.872	11.309	11.421	10.981
Below best buy limit order %	-	40.116	37.588	36.410	39.218
Order cancellation ratio %	-	75.759	72.403	68.491	65.988
Transactions					
Share of Limit Orders					
Slow speculator %	66.157	74.731	81.265	84.376	-
Investor 2 ticks PV %	52.286	52.301	51.537	50.735	46.804
Investor 4 ticks PV %	30.875	32.548	33.865	35.169	39.147
HFT %	-	33.756	47.012	56.990	65.109
Execution Times					
Slow speculator	7.774	6.394	4.939	3.918	-
Investor 2 ticks PV	5.029	4.324	3.649	2.752	1.914
Investor 4 ticks PV	3.789	3.526	3.278	3.174	2.604
HFT	-	3.188	3.746	4.142	3.556

This table presents the market quality indicators, welfare, trader strategies and transactions from model simulations performed for five different parameterizations in which the share of HFTs among speculators is progressively increased from 0% to 100%. Overall market activity and the total share of speculators in the trader population are held constant at 30%. To better compare strategies across trader groups, we report strategies for positive private valuation investors. Strategies for negative private value traders are symmetric.

Table 3.2
Impact of HFT on market quality, welfare, and strategies when adding HFTs to the trader population under symmetric information.

Share of HFTs in the Market	15%	17.5%	20%	22.5%	25%
Market Quality					
Best Bid/Offer size	1.976	2.162	2.637	3.261	4.281
Depth off BBO	5.822	7.119	8.338	9.909	11.576
Quoted spread	1.615	1.561	1.503	1.469	1.467
Effective spread	1.176	1.078	1.057	1.024	1.029
Price discovery $\lvert p_t - v_t \rvert / v_t$ %	1.289	1.185	1.138	1.101	1.118
Microstructure noise volatility	0.080	0.077	0.076	0.075	0.076
Welfare					
Average welfare	0.239	0.235	0.228	0.220	0.213
Slow speculator	0.050	0.034	0.023	0.011	0.011
Investor 2 ticks PV	0.219	0.228	0.229	0.230	0.228
Investor 4 ticks PV	0.435	0.437	0.440	0.442	0.444
HFT	0.096	0.085	0.074	0.069	0.062
Trader Strategies					
Slow speculator					
Buy market order %	0.064	0.043	0.034	0.025	0.025
Aggressive buy limit order %	1.454	1.291	1.099	0.705	0.707
At quote buy limit order %	10.779	10.174	9.684	9.966	10.013
Below best buy limit order %	36.537	37.885	37.962	40.220	40.380
Investor 2 ticks private valuation					
Buy market order %	5.420	6.203	6.463	6.846	6.859
Aggressive buy limit order %	13.777	18.525	17.020	14.812	11.870
At quote buy limit order %	39.156	23.535	23.974	23.650	30.606
Below best buy limit order %	36.161	44.639	45.313	47.052	44.746
Investor 4 ticks private valuation					
Buy market order %	16.951	16.522	14.384	15.300	15.359
Aggressive buy limit order %	17.898	18.662	17.164	16.664	19.036
At quote buy limit order %	53.357	50.203	46.913	46.831	39.505
Below best buy limit order %	10.880	14.361	17.811	20.744	19.995
High-frequency trader					
Buy market order %	0.047	0.035	0.029	0.023	0.021
Aggressive buy limit order %	1.106	1.144	0.855	0.449	0.425
At quote buy limit order %	11.309	9.568	10.471	11.742	12.881
Below best buy limit order %	37.588	38.910	38.285	37.883	35.807
Order cancellation ratio %	72.403	69.300	60.705	52.054	36.435
Transactions					
Share of Limit Orders					
Slow speculator %	81.265	85.839	87.097	90.778	90.212
Investor 2 ticks PV %	51.537	45.541	42.803	40.411	38.702
Investor 4 ticks PV %	33.865	37.331	38.214	37.085	35.739
HFT %	47.012	50.984	53.750	56.382	59.944
Execution Times					
Slow speculator	4.939	3.902	3.632	3.904	4.701
Investor 2 ticks PV	3.649	2.041	2.018	2.046	2.674
Investor 4 ticks PV	3.278	3.142	3.037	2.830	2.473
HFT	3.746	2.624	3.009	4.077	4.216

This table presents the market quality indicators, welfare, trader strategies and transactions from model simulations performed for five different parameterizations in which the share of HFTs in the overall trader population is progressively increased from 15% to 25% by gradually adding HFTs. Overall market activity increases in this scenario. To better compare strategies across trader groups, we report strategies for positive private valuation investors. Strategies for negative private value traders are symmetric.

Table 3.3
Impact of HFT on market quality, welfare, and strategies when the overall trader population is constant under asymmetric information.

Share of HFTs among Speculators	0%	25%	50%	75%	100%
Market Quality					
Best Bid/Offer size	1.334	1.454	1.582	1.837	2.441
Depth off BBO	3.523	3.530	3.822	4.522	6.765
Quoted spread	1.697	1.878	1.780	1.683	1.619
Effective spread	1.413	1.492	1.452	1.415	1.163
Price discovery $\|p_t - v_t\|/v_t$ %	1.356	1.363	1.376	1.382	1.287
Microstructure noise volatility	0.082	0.082	0.083	0.084	0.080
Welfare					
Average welfare	0.235	0.237	0.239	0.240	0.241
Slow speculator	0.069	0.062	0.053	0.034	-
Investor 2 ticks PV	0.215	0.215	0.217	0.218	0.225
Investor 4 ticks PV	0.432	0.433	0.433	0.434	0.437
HFT	-	0.108	0.098	0.092	0.073
Trader Strategies					
Slow speculator					
Buy market order %	0.199	0.163	0.132	0.121	-
Aggressive buy limit order %	2.045	3.312	2.237	2.488	-
At quote buy limit order %	11.066	11.377	11.032	8.121	-
Below best buy limit order %	38.403	36.298	37.108	37.172	-
Investor 2 ticks private valuation					
Buy market order %	5.499	5.188	5.031	5.381	5.721
Aggressive buy limit order %	10.030	10.369	12.311	13.558	18.796
At quote buy limit order %	45.305	45.089	43.027	39.022	23.746
Below best buy limit order %	32.551	33.789	32.868	33.342	43.805
Investor 4 ticks private valuation					
Buy market order %	19.777	16.810	17.571	17.435	15.686
Aggressive buy limit order %	19.543	17.508	21.585	21.419	20.295
At quote buy limit order %	47.963	55.758	44.745	42.703	42.498
Below best buy limit order %	6.766	8.381	13.174	16.518	20.794
High-frequency trader					
Buy market order %	-	0.100	0.069	0.050	0.024
Aggressive buy limit order %	-	2.110	1.782	1.604	0.724
At quote buy limit order %	-	9.801	14.471	12.895	11.884
Below best buy limit order %	-	37.688	34.955	33.916	37.387
Order cancellation ratio %	-	81.879	74.740	72.260	69.670
Transactions					
Share of Limit Orders					
Slow speculator %	66.003	71.591	74.231	70.361	-
Investor 2 ticks PV %	52.942	53.329	53.753	53.525	47.813
Investor 4 ticks PV %	32.748	34.740	32.635	34.021	37.755
HFT %	-	28.979	50.596	58.298	65.110
Execution Times					
Slow speculator	6.546	5.053	3.968	2.461	-
Investor 2 ticks PV	5.100	4.917	4.141	3.400	1.732
Investor 4 ticks PV	3.747	3.448	2.935	2.544	2.409
HFT	-	2.631	3.448	4.227	4.400

This table presents the market quality indicators, welfare, trader strategies and transactions from model simulations performed for five different parameterizations in which the share of HFTs among speculators is progressively increased from 0% to 100%. Overall market activity and the total share of speculators in the trader population are held constant at 30%. To better compare strategies across trader groups, we report strategies for positive private valuation investors. Strategies for negative private value traders are symmetric.

Table 3.4

Impact of HFT on market quality, welfare, and strategies when adding HFTs to the trader population under asymmetric information.

Share of HFTs in the Market	15%	17.5%	20%	22.5%	25%
Market Quality					
Best Bid/Offer size	1.594	1.835	1.946	2.281	2.495
Depth off BBO	3.732	5.003	5.684	6.665	7.150
Quoted spread	1.844	1.779	1.756	1.699	1.625
Effective spread	1.530	1.439	1.356	1.332	1.261
Price discovery $\|p_t - v_t\|/v_t$ %	1.376	1.310	1.266	1.239	1.194
Microstructure noise volatility	0.084	0.081	0.082	0.081	0.080
Welfare					
Average welfare	0.250	0.242	0.235	0.228	0.221
Slow speculator	0.051	0.033	0.013	-0.001	-0.013
Investor 2 ticks PV	0.217	0.223	0.228	0.229	0.229
Investor 4 ticks PV	0.434	0.436	0.437	0.442	0.447
HFT	0.101	0.094	0.088	0.080	0.073
Trader Strategies					
Slow speculator					
Buy market order %	0.154	0.120	0.125	0.115	0.134
Aggressive buy limit order %	3.525	2.551	2.409	1.724	1.529
At quote buy limit order %	10.045	9.060	8.889	9.578	10.478
Below best buy limit order %	34.896	38.626	38.139	38.588	39.642
Investor 2 ticks private valuation					
Buy market order %	5.420	5.763	6.044	6.217	6.653
Aggressive buy limit order %	12.440	16.459	18.500	17.078	16.379
At quote buy limit order %	41.317	30.980	22.593	23.842	23.337
Below best buy limit order %	33.382	37.852	41.395	44.093	46.961
Investor 4 ticks private valuation					
Buy market order %	18.543	16.263	14.509	13.232	11.566
Aggressive buy limit order %	22.554	21.351	20.549	20.428	18.278
At quote buy limit order %	40.609	46.282	48.886	39.764	33.156
Below best buy limit order %	15.667	15.667	14.399	18.110	24.482
High-frequency trader					
Buy market order %	0.078	0.056	0.044	0.039	0.031
Aggressive buy limit order %	2.484	1.987	1.824	1.200	1.044
At quote buy limit order %	11.290	11.721	10.156	11.070	11.718
Below best buy limit order %	34.209	36.410	37.674	38.794	37.973
Order cancellation ratio %	77.457	75.674	71.732	66.213	62.043
Transactions					
Share of Limit Orders					
Slow speculator %	73.135	73.036	71.930	71.937	70.757
Investor 2 ticks PV %	54.129	49.471	45.203	43.629	41.855
Investor 4 ticks PV %	34.551	38.871	42.138	41.716	40.451
HFT %	46.788	50.580	54.623	57.075	60.599
Execution Times					
Slow speculator	3.818	2.598	2.057	2.040	2.007
Investor 2 ticks PV	3.980	2.567	1.758	1.601	1.650
Investor 4 ticks PV	2.978	2.467	2.575	2.204	1.996
HFT	3.170	2.946	2.608	2.923	2.969

This table presents the market quality indicators, welfare, trader strategies and transactions from model simulations performed for five different parameterizations in which the share of HFTs in the overall trader population is progressively increased from 15% to 25% by gradually adding HFTs. Overall market activity increases in this scenario. To better compare strategies across trader groups, we report strategies for positive private valuation investors. Strategies for negative private value traders are symmetric.

Table 3.5

Impact of regulations on market quality, welfare, and strategies under symmetric information.

Setting	Base case	Minimum resting time 1.0s	Cancellation fee $0.015	Transaction tax 0.1%	Rebates $0.01/-$0.01	Speed bump 1.0s
Market Quality						
Best Bid/Offer size	1.976	1.964	1.460	2.042	1.868	4.592
Depth off BBO	5.822	5.292	2.466	5.998	4.684	12.862
Quoted spread	1.615	1.609	1.616	1.592	1.557	1.265
Effective spread	1.176	1.303	1.428	1.172	1.275	1.065
Price discovery $\|p_t - v_t\|/v_t$ %	1.289	1.358	1.360	1.282	1.341	1.337
Microstructure noise volatility	0.080	0.082	0.082	0.080	0.081	0.083
Welfare						
Average welfare	0.239	0.234	0.219	0.237	0.234	0.233
Average welfare + proceeds	0.239	0.234	0.222	0.239	0.234	0.233
Slow speculator	0.050	0.057	0.042	0.047	0.056	0.034
Investor 2 ticks PV	0.219	0.214	0.206	0.218	0.213	0.222
Investor 4 ticks PV	0.435	0.431	0.426	0.433	0.427	0.434
HFT	0.096	0.089	0.083	0.094	0.095	0.089
Trader Strategies						
Slow speculator						
Buy market order %	0.064	0.102	0.128	0.063	0.101	0.241
Aggressive buy limit order %	1.454	0.705	0.648	1.450	1.172	0.179
At quote buy limit order %	10.779	11.699	8.197	11.137	12.739	10.388
Investor 2 ticks private valuation						
Buy market order %	5.420	5.803	5.136	5.471	5.642	7.194
Aggressive buy limit order %	13.777	10.080	6.201	14.618	9.511	7.068
At quote buy limit order %	39.156	49.321	52.062	35.261	48.603	55.778
Investor 4 ticks private valuation						
Buy market order %	16.951	20.645	17.439	17.320	17.786	30.567
Aggressive buy limit order %	17.898	17.415	12.743	19.707	15.932	14.481
At quote buy limit order %	53.357	50.763	65.568	48.191	56.785	41.407
High-frequency trader						
Buy market order %	0.047	0.073	0.087	0.047	0.066	0.082
Aggressive buy limit order %	1.106	0.294	0.236	1.181	0.975	0.127
At quote buy limit order %	11.309	16.212	16.050	10.920	12.197	13.444
Order cancellation ratio %	72.403	7.892	1.127	73.185	74.833	29.026
Transactions						
Share of Limit Orders						
Slow speculator %	81.265	75.936	73.510	81.856	78.869	87.530
Investor 2 ticks PV %	51.537	54.144	53.195	51.461	53.860	51.243
Investor 4 ticks PV %	33.865	29.475	35.975	34.024	33.671	17.593
HFT %	47.012	54.069	45.960	46.120	43.242	74.078
Execution Times						
Slow speculator	4.939	7.672	21.359	4.936	6.648	9.533
Investor 2 ticks PV	3.649	4.962	6.869	3.387	5.286	4.769
Investor 4 ticks PV	3.278	3.854	4.841	3.302	4.448	3.397
HFT	3.746	9.211	14.867	3.452	4.650	10.614

This table presents the market quality indicators, welfare, trader strategies and transactions from model simulations performed for six different parameterizations. The base case is a situation where 30% of traders are speculators, half of which are HFTs, and HFTs do not have an information advantage over other traders. The five other cases introduce a minimum resting time, cancellation fees, transaction taxes, rebate fees and speed bumps of a size reported in the respective column headers. To better compare strategies across trader groups, we report strategies for positive private valuation investors. Strategies for negative private value traders are symmetric.

Table 3.6

Impact of regulations on market quality, welfare, and strategies under asymmetric information.

Setting	Base case	Minimum resting time 1.0s	Cancellation fee $0.015	Transaction tax 0.1%	Rebates $0.01/-$0.01	Speed bump 1.0s
Market Quality						
Best Bid/Offer size	1.596	1.570	1.545	1.633	1.547	2.205
Depth off BBO	3.865	3.687	2.601	3.953	3.435	4.895
Quoted spread	1.844	1.670	1.692	1.807	2.005	1.339
Effective spread	1.530	1.463	1.542	1.453	1.939	1.332
Price discovery $\|p_t - v_t\|/v_t$ %	1.376	1.440	1.419	1.381	1.522	1.395
Microstructure noise volatility	0.084	0.085	0.084	0.083	0.087	0.086
Welfare						
Average welfare	0.239	0.233	0.221	0.237	0.236	0.237
Average welfare + proceeds	0.239	0.233	0.224	0.239	0.236	0.237
Slow speculator	0.051	0.058	0.049	0.048	0.064	0.039
Investor 2 ticks PV	0.217	0.213	0.205	0.215	0.212	0.218
Investor 4 ticks PV	0.434	0.430	0.424	0.431	0.425	0.431
HFT	0.101	0.092	0.083	0.099	0.103	0.102
Trader Strategies						
Slow speculator						
Buy market order %	0.154	0.170	0.135	0.143	0.223	0.397
Aggressive buy limit order %	3.525	1.272	0.608	3.358	3.548	0.913
At quote buy limit order %	10.045	11.798	11.315	10.412	10.537	12.191
Investor 2 ticks private valuation						
Buy market order %	5.420	5.325	4.870	5.272	4.924	6.180
Aggressive buy limit order %	12.440	10.077	6.651	11.862	9.917	7.265
At quote buy limit order %	41.317	50.117	56.761	42.962	48.141	53.765
Investor 4 ticks private valuation						
Buy market order %	18.543	19.115	17.107	17.638	15.757	37.369
Aggressive buy limit order %	22.554	18.086	13.076	21.552	17.933	10.934
At quote buy limit order %	40.609	52.817	64.357	45.604	55.863	32.865
High-frequency trader						
Buy market order %	0.078	0.097	0.083	0.071	0.111	0.160
Aggressive buy limit order %	2.484	0.434	0.208	2.553	2.360	0.460
At quote buy limit order %	11.290	18.273	16.747	11.253	12.974	14.485
Order cancellation ratio %	77.457	7.975	1.195	79.108	77.173	64.577
Transactions						
Share of Limit Orders						
Slow speculator %	73.135	70.136	69.571	73.218	62.305	79.815
Investor 2 ticks PV %	54.129	55.252	54.852	53.519	56.627	56.939
Investor 4 ticks PV %	34.551	31.122	35.422	35.261	42.767	13.417
HFT %	46.788	53.369	46.644	46.803	34.623	74.599
Execution Times						
Slow speculator	3.968	12.585	21.244	4.936	4.055	10.144
Investor 2 ticks PV	4.141	6.716	9.108	3.387	8.908	3.919
Investor 4 ticks PV	2.978	3.376	4.393	3.211	3.015	3.488
HFT	3.170	8.806	14.522	3.075	3.786	6.228

This table presents the market quality indicators, welfare, trader strategies and transactions from model simulations performed for six different parameterizations. The base case is a situation where 30% of traders are speculators, half of which are HFTs. The five other cases introduce a minimum resting time, cancellation fees, transaction taxes, rebate fees and speed bumps of a size reported in the respective column headers. Throughout, HFTs have an information advantage over other traders, which only observe the fundamental value with a lag of two time units. To better compare strategies across trader groups, we report strategies for positive private valuation investors. Strategies for negative private value traders are symmetric.

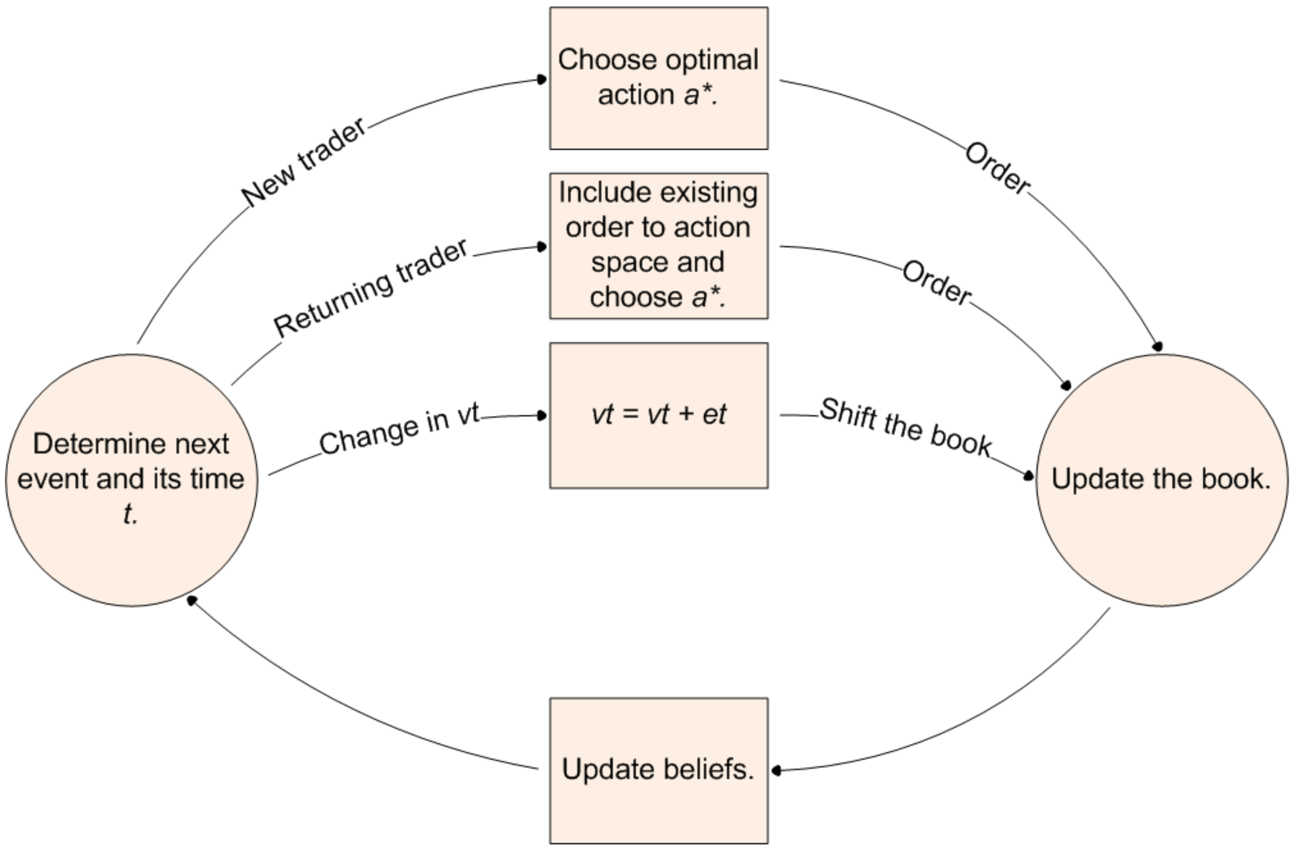

Figure 3.1: Workflow of the game.

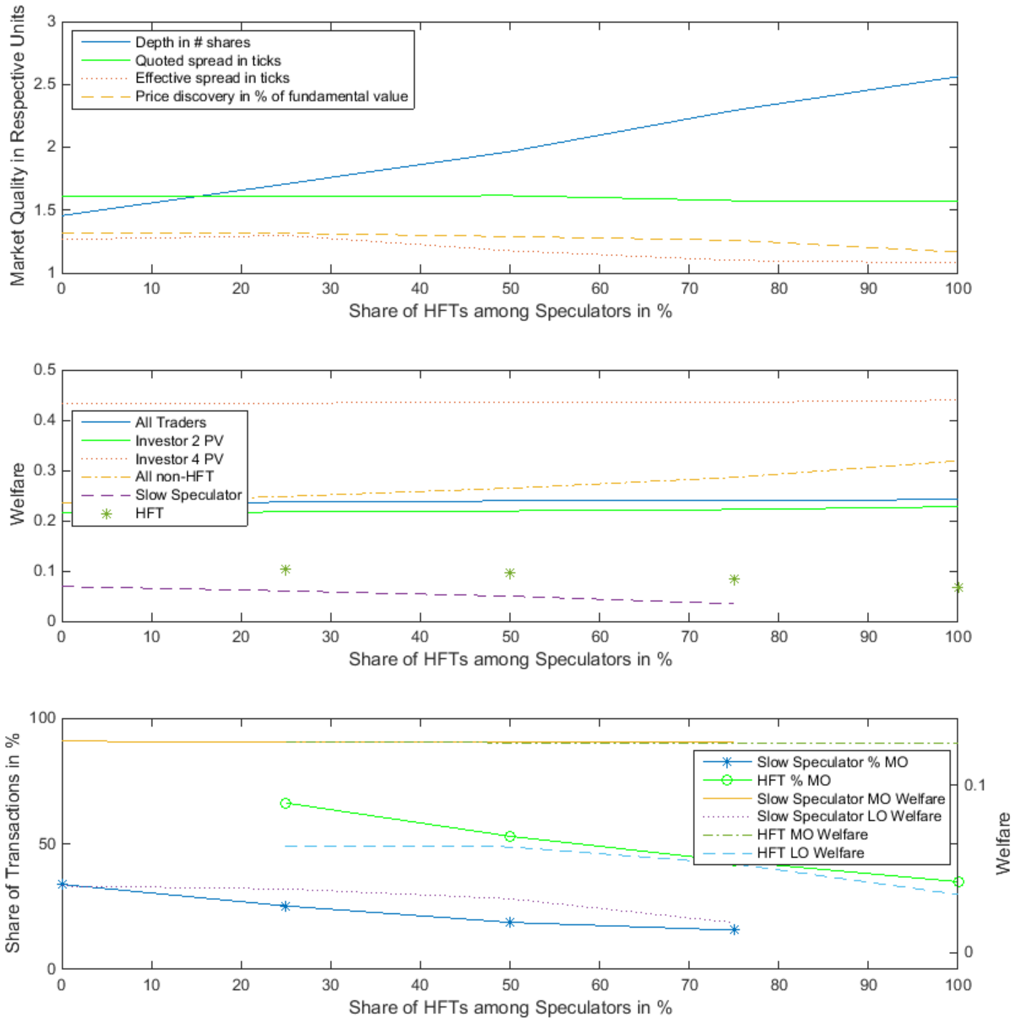

Figure 3.2: Impact of HFT on market quality, welfare, and strategies under symmetric information when the overall trader population is constant.

This figure presents the market quality indicators, welfare, and trader strategies from model simulations performed for five different parameterizations in which the share of HFTs among speculators is progressively increased from 0% to 100%. Overall market activity and the total share of speculators in the trader population are held constant at 30%.
The top panel reports the different market quality measures as a function of the share of speculators in the market that are HFTs. The middle panel reports welfare for the different groups and the average welfare of all traders in the market. The bottom panel reports the share of market orders in slow speculators' and HFTs' transactions (left axis) and their welfare from market orders and limit orders (right axis).

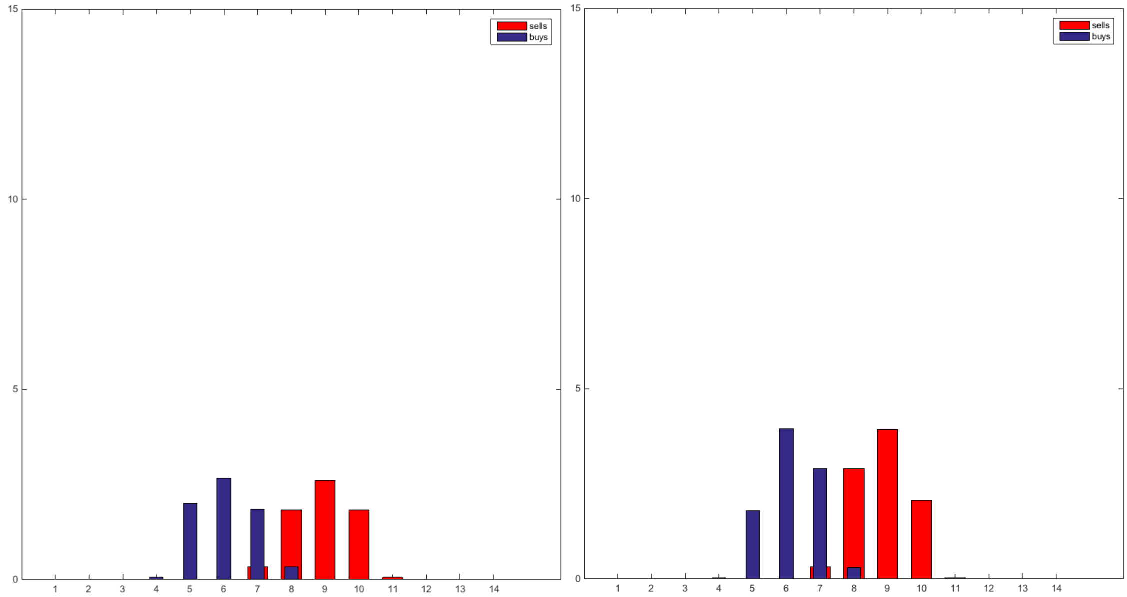

Figure 3.3: Average composition of the limit order book when replacing slow speculators by HFTs.

Average composition of the limit order book from model simulations performed for two different parameterizations in which the share of HFTs among speculators is increased from 25% (left) to 75% (right). Overall market activity and the total share of speculators in the trader population are held constant at 30%.

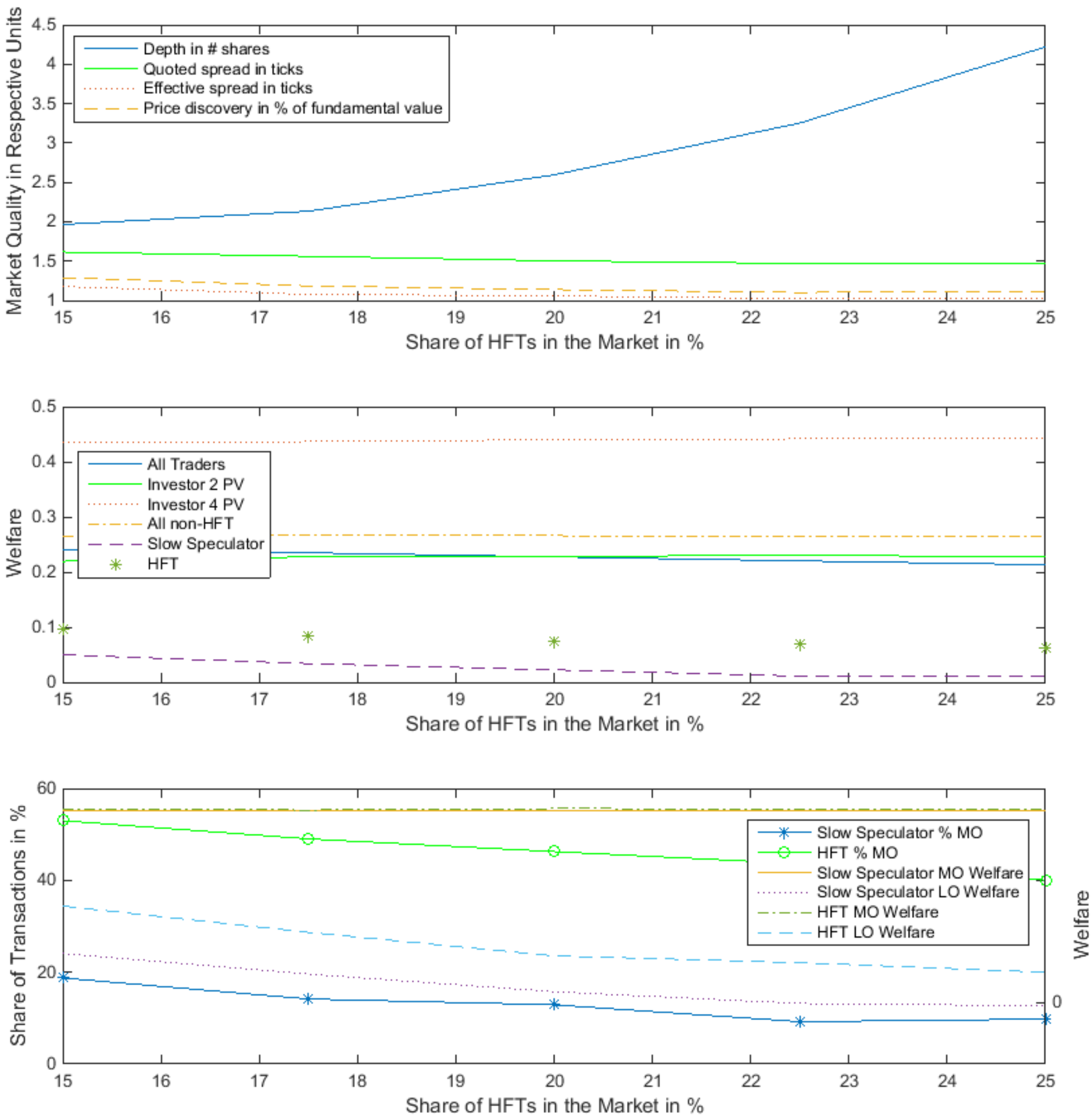

Figure 3.4: Impact of HFT on market quality, welfare, and strategies under symmetric information when adding HFTs to the trader population.

This figure presents the market quality indicators, welfare, and trader strategies from model simulations performed for five different parameterizations in which the share of HFTs in the overall trader population is progressively increased from 15% to 25% by gradually adding HFTs. Overall market activity increases in this scenario.
The top panel reports the different market quality measures as a function of the share of traders in the market that are HFTs. The middle panel reports welfare for the different groups and the average welfare of all traders in the market. The bottom panel reports the share of market orders in slow speculators' and HFTs' transactions (left axis) and their welfare from market orders and limit orders (right axis).

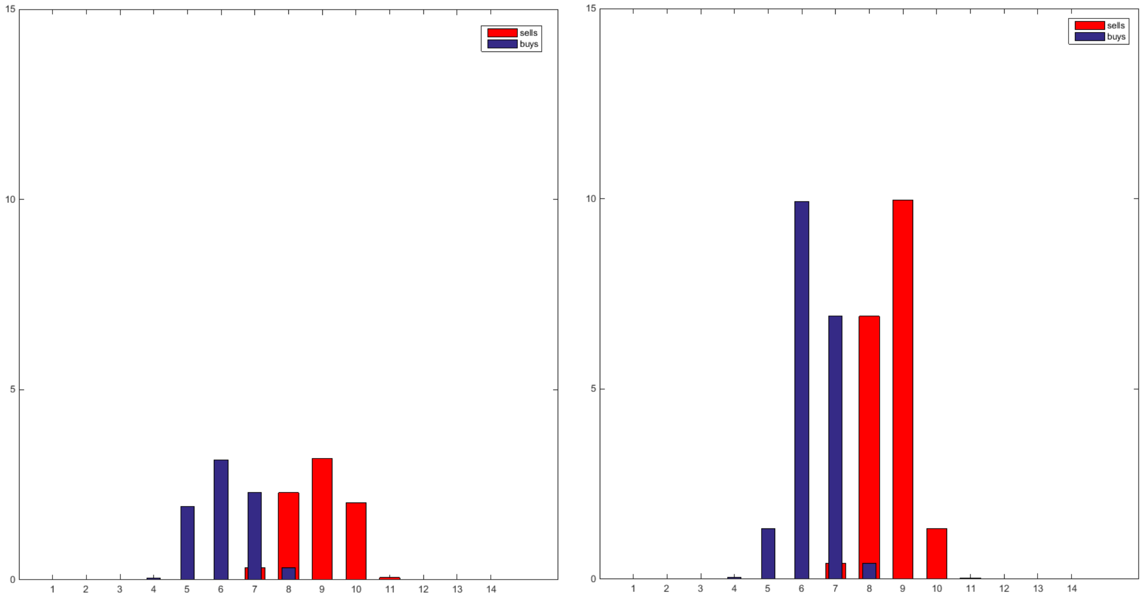

Figure 3.5: Average composition of the limit order book when adding HFTs.

Average composition of the limit order book from model simulations performed for two different parameterizations in which the share of HFTs in the overall trader population is increased from 15% (left) to 25% (right) by adding HFTs. Overall market activity increases in this scenario.

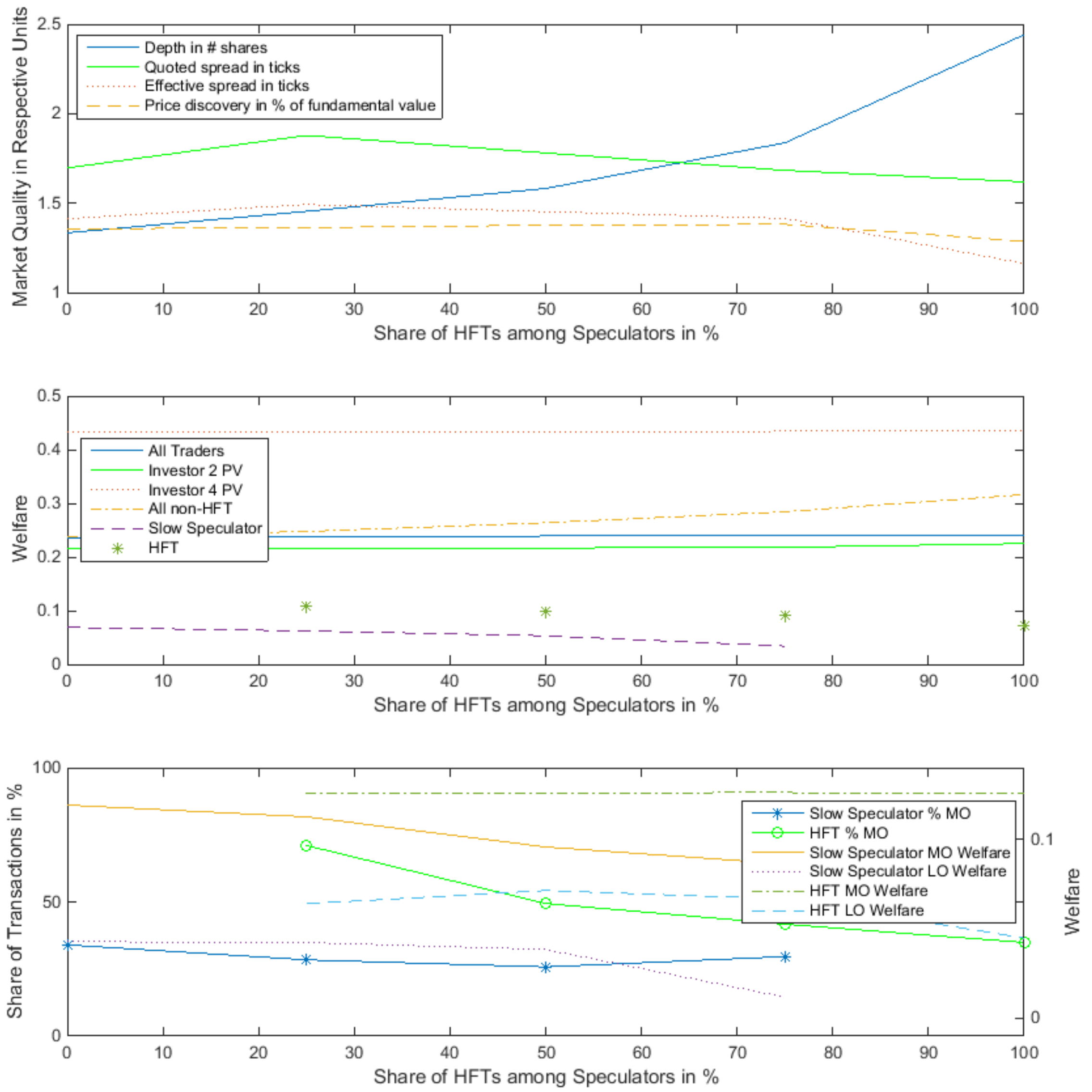

Figure 3.6: Impact of HFT on market quality, welfare, and strategies under asymmetric information when the overall trader population is constant.

This figure presents the market quality indicators, welfare, and trader strategies from model simulations performed for five different parameterizations in which the share of HFTs among speculators is progressively increased from 0% to 100%. Overall market activity and the total share of speculators in the trader population are held constant at 30%.
The top panel reports the different market quality measures as a function of the share of speculators in the market that are HFTs. The middle panel reports welfare for the different groups and the average welfare of all traders in the market. The bottom panel reports the share of market orders in slow speculators' and HFTs' transactions (left axis) and their welfare from market orders and limit orders (right axis).

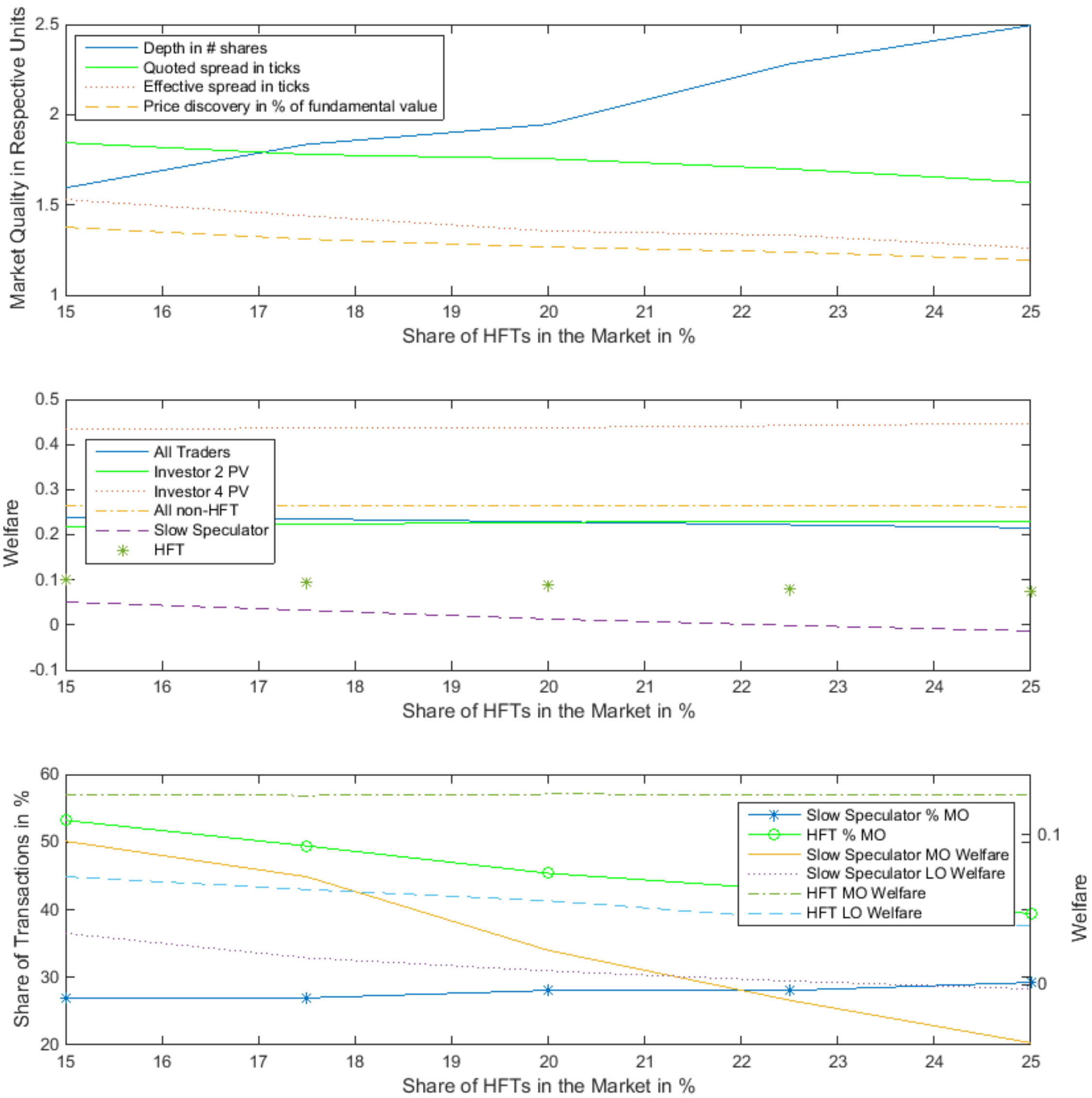

Figure 3.7: Impact of HFT on market quality, welfare, and strategies under asymmetric information when adding HFTs to the trader population.

This figure presents the market quality indicators, welfare, and trader strategies from model simulations performed for five different parameterizations in which the share of HFTs in the overall trader population is progressively increased from 15% to 25% by gradually adding HFTs. Overall market activity increases in this scenario.
The top panel reports the different market quality measures as a function of the share of traders in the market that are HFTs. The middle panel reports welfare for the different groups and the average welfare of all traders in the market. The bottom panel reports the share of market orders in slow speculators' and HFTs' transactions (left axis) and their welfare from market orders and limit orders (right axis).

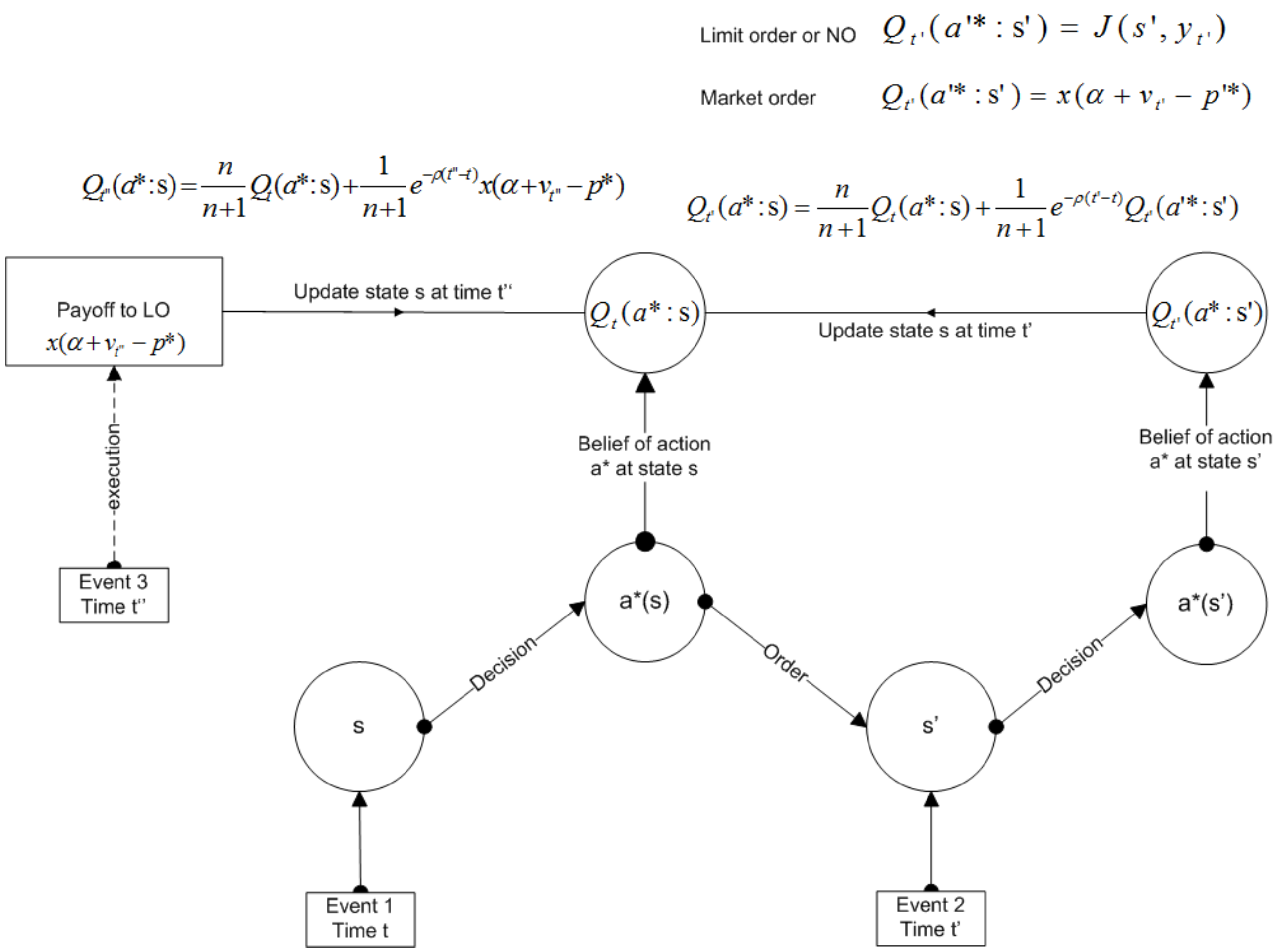

Figure 3.8: Updates of beliefs

At each time t in each state s, each action $\tilde{a}$ has an associated payoff $Q_t(\tilde{a}|s)$. The beliefs at each time t imply an optimal strategy y_t, which assigns the payoff maximizing action at each state, $a^*(s) \in \arg\max_{\tilde{a}\in\mathcal{A}(s)} Q_t(\tilde{a}|s)$. The value of state s then is $J(s, y_t) = Q_t(a^*(s)|s)$. If the previous state was s and the state s' is hit, the continuation value $J(s'|y_{t'})$ is based on the action a' taken in the new state s':

1. *Market order*: payoff from the market order.
2. *Limit order*: expected value of the limit order represented by the action a', $Q(a'|s')$.
3. *No order*: expected value of taking no action a', $Q(a'|s')$.

Put together, the Q-factors are updated in the following fashion, where $J(s'|y_{t'})$ represents the payoff to the action taken at state s' and time t'.

$$Q_{t'}(a^*|s) = \frac{n}{n+1} Q_t(a^*|s) + \frac{1}{n+1} e^{-\rho(t'-t)} J(s'|y_{t'})$$

Similarly, if the previously submitted limit order is executed, the expected payoff at the previous state is updated by $\tilde{x}(\alpha + v_{t'} - \tilde{a}^*)$, where t' is the current time.

Chapter 4

Price Impact of Aggressive Liquidity Provision

Ramazan Gençay, Soheil Mahmoodzadeh, Jakub Rojček and Michael C. Tseng[1]

Abstract

This paper analyzes brief episodes of high-intensity quotes turnover and revision—"bursts" in quotes—in the U.S. equity market. Such events occur very frequently, around 400 times a day for actively traded stocks. We find significant price impact associated with this market-maker initiated event, about five times higher than during non-burst periods. Bursts in quotes are concurrent with short-lived structural breaks in the informational relationship between market makers and market takers. During bursts, market makers no longer passively impound information from order flow into quotes—a departure from the traditional market microstructure paradigm. Rather, market makers significantly impact prices during bursts in quotes. Further analysis shows that there is asymmetry in adverse selection between the bid and ask sides of the limit order book and only a sub-population of market makers enjoys an informational advantage during bursts. Our results call attention to the need for a new microstructure perspective in understanding modern high-frequency limit order book markets.

Keywords: Price Impact, Burst, High-Frequency Trading, Market Quality, Adverse Selection

JEL classification: G14 · G28 · C58.

4.1. Introduction

The increasing prevalence of algorithmic trading has spurred a considerable amount of work seeking to understand the impact of high-frequency trading on market quality and its influence on execution

[1]Ramazan Gençay gratefully acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada and the Social Sciences and Humanities Research Council of Canada. Jakub Rojček gratefully acknowledges the support of the Swiss Finance Institute's Advanced Doctoral Student Grant during his visit to Simon Fraser University. We are grateful for fruitful discussions, comments and suggestions made by Michel Habib, Joel Hasbrouck, Hubert Kempf, Marc Paolella, Angelo Ranaldo, Thomas Richter, Jean-Charles Rochet, Alexander Wagner, Alexandre Ziegler, and seminar participants at the University of Zurich. All errors and omissions remain ours.

quality for traditional institutional investors.[2] As pointed out by O'Hara (2015), the rapid proliferation and evolution of high-frequency technology presents challenges to the existing market microstructure paradigm. While theoretical models have shown that algorithmic traders in continuous limit order markets derive competitive advantage from faster analysis of order book evolution and other information,[3] the empirical literature has focused on the impact of algorithmic trading on traditional measures of market quality.[4] Armed with the ability to update and revise quotes at an extremely fast rate, the behavior of high-frequency market makers may depart fundamentally from that assumed in the market microstructure literature. This issue remains largely understudied. In this paper, we aim to fill this gap by analyzing market makers' activity that manifests itself directly in the quotation process and is not immediately attributable to the traditional role of passive liquidity provider.

Understanding the various dimensions of the quotation process in financial markets is crucial, as it spans and interweaves different market participants, such as market makers, speculators, institutional investors and regulators, and different objectives, such as inventory management, optimal execution, arbitrage, price discovery, and fairness of the trading process. One aspect of the quotation process that is of particular concern to regulators is irregularity in arrival intensity. Regulatory questions regarding the impact of high-frequency traders' ability to manipulate the quotation process on the market remain unanswered.[5]

While unrelated with the fundamental value of assets, we show in this paper that market-maker initiated irregularity in the quotation process, nevertheless, has price impact. Many exchanges have rushed to introduce rules designed to minimize such, apparently predatory, market making activity. For example, traders may be penalized for updating their quotes more than 100 times per transaction, by so-called *cancellation fees*.[6] The empirical analysis of new regulations thus comes *ex post*.[7] Adopting *ad hoc* regulations in response to a poorly understood phenomenon may lead to suboptimal market

[2] For an overview of the literature on high-frequency trading, see the surveys by Jones (2013) and O'Hara (2015).
[3] Foucault et al. (2015), Aït-Sahalia and Saglam (2014), Biais et al. (2015) and Hoffmann (2014).
[4] See Jovanovic and Menkveld (2011), Brogaard (2010), Hasbrouck and Saar (2013), Riordan and Storkeimeier (2012), Tong (2013) and Hendershott et al. (2011).
[5] These concerns are stated in SEC (2010) and CESR (2010), for example.
[6] Examples of regulatory initiatives can be found in Aït-Sahalia and Saglam (2014), Colliard and Hoffmann (2013), Friedrich and Payne (2015), and Rojcek and Ziegler (2015).
[7] The experimental approach in regulation seems to have become a new feasible way to test new regulatory proposals, supported and initiated also by the Securities and Exchange Commission (SEC) in the recent controlled tick size change event, see SEC (2014).

outcomes.[8] In this paper, we provide a detailed analysis of market quality during periods where the quotation process exhibits extreme behavior. We show that during very brief periods of aggressive liquidity provision, *bursts in quotes*, market makers are able to anticipate—possibly precipitate—price impact and trade in the direction of the subsequent price change.[9] Aggressive liquidity provision, in contrast to passively absorbing incoming market orders, is a form of information inflow into the market. Further analysis on this departure from the traditional microstructure paradigm shows that asymmetry in adverse selection exists during, and only during, bursts in quotes between market makers on opposite sides of the limit order book. Market makers on the side opposite of the burst suffer elevated adverse selection costs, while market makers on the side of the burst enjoy negative price impact.

The rest of the paper is organized as follows. Section 4.2 surveys the related literature. Section 4.3 describes the data selection, and Section 4.4 explains the definition and detection of bursts. Section 4.5 gives a detailed analysis of market quality during bursts. Section 4.6 states our main results on the price impact of bursts in quotes. Section 4.7 refines the analysis by examining the asymmetry in adverse selection faced by market makers on opposite sides of order book during bursts in quotes. Section 4.8 concludes.

4.2. Literature Review

Our paper contributes to the growing literature analyzing the role of quotations in current financial markets that are dominated by algorithmic and high-frequency traders using limit order book as the underlying matching mechanism. The rise in quote cancellations in the millisecond environment is documented, e.g. by Hasbrouck and Saar (2013) and many others. A few streams of thought are present within this domain. Hasbrouck (2015) observes that at sub-second horizons bids and asks in US equity markets are more volatile than implied by long-term fundamentals. In considering underlying causes, Hasbrouck (2015) suggests that this volatility is not likely to arise from quote-stuffing, spoofing, or mixed-strategy behavior, but more likely reflects Edgeworth cycles, or recurrent phases of undercutting. This comports with Baruch and Glosten (2013), who show that repeatedly playing a certain static liquidity provision game has a subgame perfect equilibrium in mixed strategies, which correspond

[8] For example, if exchanges or regulators considered imposing a cap on the order to trade ratio or rate of quote arrivals, we suggest first analyzing liquidity, price impact, and other market quality measures precisely during periods where the quotation process exhibits extreme behavior.

[9] In our sample, the duration of such periods of extremely high order arrival intensity is approximately *1.09* seconds, implying that high-frequency traders are the primary liquidity suppliers during such events.

to fleeting orders and flickering quotes.[10] The natural question is then under what conditions is the market in an equilibrium with flickering quotes and what the corresponding conditions for the quotation process are. As we show in this paper, extreme quotation activity only occurs in brief and irregular episodes.

Other studies, somewhat similar to ours, have focused their attention on analyzing the number of quotes involved in the trading process. Egginton et al. (2014) find that stocks experience decreased liquidity, higher trading costs, and increased short term volatility during periods of intense quoting activity. We confirm these findings in Section 4.5 and further examine the issue of permanent price impact and adverse selection during bursts in quotes. On the order flow side, Conrad et al. (2015) explores market resiliency during periods of exceptionally high-intensity and low-latency trading—large liquidity drawdowns in which, during a millisecond, trading algorithms systematically sweep large volume across multiple trading venues using predominantly intermarket sweep orders (ISO's).[11] Although such large drawdowns incur trading costs, they do not appear to degrade the subsequent price formation process or increase the subsequent cost of trading. Chakravarty et al. (2012) analyze the role of ISO's. They find that ISO trades have a significantly larger information share despite their small trade size relative to non-ISO trades. We also find that during bursts in quotes, compared to periods with no bursts in quotes, the share of ISO trades rises from roughly 39% to 49% in our sample. On the price process, Brogaard et al. (2015) find that, during extreme price movements, high-frequency traders act as net liquidity suppliers, while non-high-frequency traders act as net liquidity demanders. Moreover, high-frequency traders are active liquidity providers during price jumps that result in permanent price changes, absorbing the most informed order flow. Our results verify a conditional converse: during periods of burst in quotation activity, when liquidity suppliers are necessarily high-frequency traders, volatility is significantly higher, and high-frequency market makers on the opposite side of the burst absorb incoming order flow that is more informed than during non-burst periods. This finding indicates that part of the order flow consumed during extreme price movements necessarily includes slower market makers being picking off. We also find that the burst itself has price impact and it magnifies the price impact of those trades which are on the opposite side of the burst. Utilizing the NASDAQ dataset which identifies high-frequency traders, Brogaard et al. (2014) shows that high-

[10] The notion of flickering quotes corresponds to bursts in quotation activity in this paper.

[11] ISO's are marketable limit orders designed to be sent to multiple venues simultaneously. They are meant to facilitate Reg NMS' basic principle of routing orders to the exchange claiming the NBBO.

frequency trading is correlated with public information and that high-frequency traders trade in the direction of permanent price impact and opposite of transitory pricing errors. Our results are different in that, in addition to competition for the order flow, bursts in quotes may be the outcome of a zero-sum game between market makers. Brogaard et al. (2015) considers the price impact of limit orders using Canadian regulatory data on a trade by trade basis. They find that high-frequency traders are responsible for the biggest portion of price discovery through their limit orders. Our finding can be interpreted in the opposite direction. Rather than correcting price mis-alignment, it is possible that bursts in quotes are used to introduce price mis-alignment.

Our paper also relates to the literature on the price impact of trades, with the additional feature of incorporating bursts in quotes, which have not been analyzed previously. Building on the permanent price impact model of Hasbrouck (1991), several models incorporating trading activity and its non-linear impacts have been put forward (see e.g. Dufour and Engle (2000), Zhang et al. (2001), Engle and Lunde (2003), and Hautsch and Huang (2012)).[12] Related to our paper is the paper by Cont et al. (2014), who enhance the traditional definition of the order flow by the no-trade limit order book events and show that such an approach dramatically increases the explanatory power of the basic price impact regression. While supporting this finding by showing that bursts in quotes have significant price impact, we also explore information asymmetry in the market.[13] In addition, we consider the economic implications of bursts taking place on the same and opposite side of the book as order flow.

Various regulators and self-regulatory organizations have introduced or are contemplating introducing caps on quote-to-trade ratios. The obvious implementation is introducing cancellation fees. Cancellation fees have been introduced by the Canadian regulatory authority IIROC with the explicit aim of curbing high-frequency spam. The Milan Stock Exchange and the NASDAQ Stock Market introduced an excess order fee for high quote-to-trade ratios in April and July 2012 (see SEC (2012)). Cancellation fees are also imposed by Eurex for order-to-trade ratios exceeding 5, on NASDAQ OMX for ratios exceeding 100, as well as on the German stock exchange. The new MiFID2 European Directive also foresees the introduction of cancellation fees to curb excessive quote-to-trade ratios. Empirical evidence on the impact of cancellation fees is provided by Malinova et al. (2013), who find that following their introduction on the Toronto Stock Exchange, quoting activity decreased by 30% and spreads

[12] Further useful surveys on the empirical estimation of price impact are provided in Bouchaud (2010) and Pham et al. (2015).

[13] Cont et al. (2014) explicitly eschews the adverse selection perspective.

increased by 9%. Friederich and Payne (2015) study the cancellation fees introduced above a certain quotes-to-trades ratio threshold on the Italian Stock exchange and observe a decrease in market depth and worsened price discovery. Our results indicate two possible channels why curbing the amount of quotes might be detrimental to market quality. First, such penalties remove an information channel to the market, since quotes have information content. Second, bursts in quotes on the same side accommodate market taker's order flow with a much smaller increase in effective spreads.

4.3. Data Description

Data for the analysis are taken from the Trades and Quotes (TAQ) database, which is a collection of intraday trades and quotes for all securities listed on the New York Stock Exchange, American Stock Exchange, Nasdaq National Market System and SmallCap issues, which are traded on 15 major U.S. exchanges within the U.S. National Market System (NMS). Quotes data consists of limit orders eligible for the National Best Bid or Offer (NBBO) and collected through the Securities Information Processor (SIP). TAQ does not contain quotes at deeper levels of the book. Transactions data consists of all transactions carried out within the NMS. Two sets of data are used. For sampling frequencies of one second and five seconds, we use data of 15 very active and liquid companies during the approximately four months period from October 10, 2011 to February 7, 2012.[14] The summary statistics in terms of average number of trades, quotes, price, lot size, and quoted spread are shown in Table 4.1. For a sampling frequency of 100 milliseconds, more recent, millisecond time-stamped TAQ data is used. Based on the activity criterion, we choose a sample of 20 active stocks for the period from March 9, 2015 through April 8, 2015.[15] We divide trading days into equidistant time intervals of 100 milliseconds, one second, and five seconds. One trading day is six hours from 9:45 until 15:45, disregarding the first and last 15 minutes of the official trading hours for possible excessive volatility. We collect a number of variables for each interval—the number of changes in the NBBO across exchanges, the number of changes of the National Best Offer across exchanges, the number of changes of the National Best Bid across exchanges, NBBO quoted prices, quoted mid-point, and average quoted spread. We point out that the changes to the NBBO may be due to a number of possible scenarios. An existing quote may

[14] Out of the 30 highest capitalization stocks traded on US exchanges, we selected those 15 which had the highest daily average volume.

[15] The 20 stocks consist of 10 stocks with the highest average daily traded volume during March 2015 on NASDAQ and 10 stocks based on the same criterion on NYSE.

be revised on price or size. A new better quote may be transmitted to the NMS. The existing best quote may be cancelled causing the second best quote to become the new best quote. An incoming transaction may be executed against the best quote or quotes, in which case the new best quote is determined from the remaining active quotes in the NMS. Regulations require that transactions are reported with a maximum delay of 90 seconds.[16]

Transactions which happen in the upstairs market or dark pools might take place at prices that differ significantly from those observed on lit exchanges or they can be reported with a delay. Transactions which happen outside of the prevailing bid and ask quotes are known as "out-trades." Out-trades would generate bouncebacks in the transaction price process[17], which might compromise our volatility or effective spread calculations. To filter out transactions which would introduce bouncebacks due to out-trades reported with a delay possibly longer than 90 seconds, or negotiated at prices outside of the prevailing quotes, we use the cleaning procedure described in Aït-Sahalia and Jacod (2014). We discard transactions at prices which are outside the tunnel constructed by using the worst past 90-seconds trailing NBBOs and deem the remaining transactions as eligible. Specifically, first we compute the average NBBO for each interval. Next, for each interval, we construct a range, which is formed by the minimum average bid and maximum average ask during the past 90 seconds. To determine whether a transaction which happened during the particular interval is eligible, we test if the transaction falls within this range. Afterwards, we collect the number of transactions and volume weighted average price (VWAP).[18] Using the methodology of Lee and Ready (1991), we further assign a buy or sell sign to obtain the order flow variable.[19]

4.4. Bursts in Financial Data

"One man's noise is another man's signal."

As described in Section 4.3, limit order book statistics of interest are computed for subintervals at a given frequency—e.g. the number of quote revisions for every second. We then define *bursts* to be

[16] Exchanges report transactions continuously. The maximum delay restriction applies to transactions from dark pools and the upstairs market.

[17] For example an out-trade transaction happening at a price above the prevailing ask will result in a positive return, followed by a negative return.

[18] VWAP is computed by summing up the dollar volume of all transactions in an interval and dividing by the total number of shares traded in the interval.

[19] The same eligibility criterion and trade-direction assignment procedure is applied to data with a 100-milliseconds sampling frequency.

those subintervals during which the given statistic has extremely large realizations—e.g. when quote revisions occur at an extremely high rate per second. We define "extremely large" by choosing a tail probability threshold—e.g. $Prob = 99.9\%$. It turns out that the unconditional distributions of financial time series we consider in this paper fit well the exponential distribution (see also Vlachos et al. (2008)). We therefore choose the threshold for bursts corresponding to tail probability $Prob$ as $x = -\hat{\mu}\log(1 - Prob)$, where $\hat{\mu}$ is the sample mean of the quotes revision time series during a corresponding interval.[20]

The limit order book statistics considered are the number of transaction arrivals, total quote revisions on both sides of the book, and quote revisions on the bid and ask sides separately. Bursts in our sample are therefore determined in two steps. First, the threshold for a chosen probability is computed for every 15-minute interval for each statistic using the formula $x = -\hat{\mu}\log(1 - Prob)$. Each 15-minute interval thus has its own specific threshold. To mitigate the in-sample impact on the threshold of possible extreme behavior of the considered times series, we include the 5 minutes before and after the corresponding interval in estimating the mean of the distribution, μ. An example of thresholds computed for two stocks for quote revisions and transaction arrivals within the same 15-minute interval is shown in Figure 4.1. These thresholds are then applied to equidistant sub-partitions of the 15-minute intervals for quotes and trades. The summary statistics for one trading day using $Prob = 99.9\%$ and subintervals of 100 milliseconds, one second, and five seconds for measuring bursts in quotes are shown in Table 4.2. As expected, we observe that each summary statistic varies monotonically with the sampling frequency. Higher sampling frequency—corresponding to shorter subintervals—correspond to a higher number of bursts, shorter average duration of bursts, shorter inter-arrival time between bursts, higher volume traded during bursts, and total longer duration of bursts. We adopt the sampling frequency of one second for the subsequent analysis. The threshold for the number of quotes and trades for one trading day at one-second frequency in Alcoa (AA) stock is shown in Figure 4.2.

[20]Given the exponential assumption, our definition of burst coincides with that of *unified outlier* in Knorr and Ng (1997). Alternatively, one may simply choose the threshold to be the corresponding quantile of the empirical distribution. Both definitions yield similar empirical results.

4.5. Impact of Bursts on Market Quality

As the initial step in our analysis, we evaluate the impact of bursts in quotes on basic market quality measures, whose definitions are recalled below. In Section 4.7, we further consider measures of realized spread and adverse selection.

- The *quoted spread* QS_t during the t-th interval is measured as the average difference between the ask price A_t and bid price B_t for the given interval, $QS_t = \overline{A_t - B_t}$.
- The *effective spread* during the t-th interval is defined as $ES_t = 2q_t(P_t - M_t)$, where P_t is the volume weighted average price (VWAP) in the t-th interval, M_t is the average mid-quote from the t-th interval, and q_t is the trade direction indicator.[21] The effective spread measures market maker's revenue for supplying immediacy.[22]
- The *return volatility* is defined as $\sigma_t = \frac{|P_t - P_{t-1}|}{M_{t-1}}$.[23]

As stated in Section 4.4, inspecting the summary statistics reported in Table 4.2 leads us to use a sampling frequency of one second. To reinforce this point, in this section we also report the market quality measures computed based on higher (100-milliseconds) and lower (5-seconds) frequencies.[24]

Figure 4.3, Figure 4.4 and Figure 4.5 examine patterns in the quoted spread, the effective spread, and return volatility in 2.5-second, 25-second and 2-minute windows centered around bursts in quotes. The panels, from left to right, show market quality measures computed at 100-millisecond, one-second, and five-second sampling frequencies, respectively. For example, for the 100-millisecond sampling frequency (left panel), 100-millisecond sub-interval averages are computed across 2.5-second windows centered around the burst, across all stocks in our sample. The initial values for each burst event are normalized to 100. The sample of each market quality measure is displayed as the black solid line. The shaded area represents one standard deviation from the mean. Corresponding time series for the one-second

[21] $q_t = 1$ if net aggregate order flow during the t-interval amounts to a buy order and $q_t = -1$ if net aggregate order flow during the t-interval amounts to a sell order. To compute q_t, we first determine the trade direction indicator for each transaction that occurs during the interval using the Lee and Ready (1991) criterion, namely $+1$ if buyer initiated and -1 if seller-initiated. The trade direction is then assigned to the volume (number of share traded) in that transaction. The total signed volume defines the order flow variable O_t. This variable may be positive or negative depending on whether more shares were bought or sold, and $q_t = \text{sgn}(O_t)$.

[22] Equivalently, the effective spread measures market takers' execution cost.

[23] Our definition is close to that for the realized volatility of the process $\log P_t$ where P_t is modelled as a semimartingale.

[24] The main part of our analysis in Section 4.6 and Section 4.7 will be based on one-second frequency.

and five-second intervals are displayed in the middle and right panels. We also show the average order size during bursts in quotes in Figure 4.6.

Figure 4.3 shows that a five-second sampling frequency is unable to detect the spike in the quoted spread seen at higher frequencies. The noise content is also higher at the five-second frequency. Not surprisingly, the spike in quoted spread is accompanied by a concurrent spike in return volatility with similar signal-noise characteristics with respect to sampling frequency shown in Figure 4.5. Also telling is the sharp dip in average order size during bursts shown in Figure 4.6, showing a corresponding change in market taking behavior as well as market making. Average order size decreases by 80% at 100-milliseconds frequency. Lower order size and higher correlation with special order types[25] indicates increased presence of sophisticated market participants. A lower sampling frequency of five seconds again shows less pronounced patterns. The concurrent spike in effective spread shows that the market makers are able to extract higher execution cost during bursts in quotes relative to non-burst periods. In contrast to other measures, the effective spread is more noisy around bursts at higher frequencies and consolidates when measured across five-second intervals. Taken together, the market quality measures indicate that trading is more volatile during bursts in quotes. Our analysis thus far—in particular, the duration and sensitivity with respect to sampling frequency—points towards the characterization of bursts as unanticipated short-lived events whose market signature dissipates quickly. The frequency at which limit order book statistics associated to bursts are visible also suggests that these are events for which the primary—possibly the only—instigators are high-frequency traders.

4.6. Empirical Analysis of Price Impact

Price impact usually refers to measures of the information content of a trade. A trade can have a permanent and temporary price impact. The permanent price impact is impounded to asset prices, whereas the temporary price impact dissipates as time passes. In this section, we study the permanent price impact and the role that bursts in quotes play in determining permanent changes in prices.

4.6.1. Price Impact during Bursts

To examine the process with which information is incorporated into prices, we extend the VAR model of Hasbrouck (1991) to reflect burst activity. Let r_t denote the change in the market maker's estimate

[25]Correlation with the number of ISOs increases from 39% to 45% during periods of bursts in quotes.

of fundamental value and q_t be the (aggregated) order flow direction, positive for buyer-initiated trades and negative for seller-initiated trades, as defined in Section 4.5.[26] We start with the basic assumption that trades have linear price impact[27], i.e.

$$r_t = \beta_t q_t + \varepsilon_t \tag{4.1}$$

where ε_t is information available to the market maker in addition to information from order flow $\{q_t\}$. Ensuring exogeneity of ε_t and its interpretation as the market maker's information uncorrelated with order flow requires controlling for microstructure frictions. For example, inventory-control effects may cause past quote updates to influence the market maker's current quote update. Delayed adjustment to information and expectation of order-splitting on the part of the market takers means order flow may have lagged effects on quote updates. This suggests an infinite-order autoregressive specification[28]

$$r_t = \underbrace{\alpha_1 r_{t-1} + \alpha_2 r_{t-2} + \cdots}_{\text{inventory-control effects}} + \beta_0 q_t + \underbrace{\beta_1 q_{t-1} + \beta_2 q_{t-2} + \cdots}_{\text{delayed adjustment, response to order-splitting, etc}} + \underbrace{\varepsilon_t}_{\text{market maker's information}}. \tag{4.2}$$

On the market taker side, similar microstructure considerations apply to order flow $\{q_t\}$, which is therefore assumed to follow a similar specification:[29]

$$q_t = \gamma_1 r_{t-1} + \gamma_2 r_{t-2} + \cdots + \delta_1 q_{t-1} + \delta_2 q_{t-2} + \cdots + \underbrace{\nu_t}_{\text{market takers' private information}} \tag{4.3}$$

where the trade innovation ν_t is market takers' private information. Assuming all regressors are exogenous, embedded in the above VAR model of $\{r_t, q_t\}$ is the fundamental assumption that ε_t and ν_t are orthogonal.[30] This underpins the traditional microstructure perspective—there is no contemporaneous causality running from the market makers' quote adjustment to order flow. On the other hand, market takers' private information is reflected by the market maker's update to his estimate of the fundamen-

[26]We use the midpoint of the quoted spread as a proxy for the market maker's estimate as in Hasbrouck (1991). r_t is the percentage mid-quote return based on the average mid-quote immediately before t and immediately after t.

[27]Many theoretical microstructure models, spanning from the seminal Kyle (1985) model to, for example, Duffie and Zhu (2015), derive an equilibrium where the functional form of price impact of trades is linear.

[28]In estimating the model, we found that r_t has significant autocorrelation whose directions alternate with respect to lag increments, consistent with a market maker who is inventory-neutral. See Table 4.6 in the Appendix. Also as expected, past order flow, up to 10 lags, has price impact. See Table 4.7 in the Appendix.

[29]Order flow $\{q_t\}$ may exhibit auto-correlation due to order splitting. Autocorrelation in $\{q_t\}$ may also arise as an equilibrium outcome when the market taker anticipates the market maker's inventory-neutral tendency and may adjust his order flow according to past quote updates. In estimating the order flow equation, we do find the coefficients γ_t mirror the behavior of α_t in alternating signs, but with opposite signs. See Table 4.7 in the Appendix.

[30]Assume that the market makers' public information ε_t is exogenous to past returns $\{r_{t-1}, r_{t-2}, \cdots\}$ and past order flow $\{q_{t-1}, q_{t-2}, \cdots\}$. Then ε_t is exogenous to q_t if and only if it is exogenous to ν_t.

tal value of the asset. The burst activity we consider, which is a fundamentally new phenomenon of the modern high-frequency market, suggests the possibility of departure from this informational dichotomy. We also point out that Equation (4.2) and Equation (4.3) impose no structural restriction on the exact nature of information being incorporated into quote updates, except that trade information ν_t and non-trade information ε_t are orthogonal. This flexibility allows us to transport and extend the model into the high-frequency setting.[31]

To analyze the permanent price impact of bursts in quotes, define the time series indicators B_t^Q and B_t^T by[32]

$$\mathrm{B}_t^Q = \begin{cases} 1 & \text{if there is a burst in quotes in period } t \\ 0 & \text{otherwise,} \end{cases} \tag{4.4}$$

$$\mathrm{B}_t^T = \begin{cases} 1 & \text{if there is a burst in trades in period } t \\ 0 & \text{otherwise.} \end{cases} \tag{4.5}$$

Allowing for additional price impact during bursts in quotes and trades leads to the specification

$$r_t = \sum_{i=1}^{\infty} \alpha_i r_{t-i} + \sum_{i=0}^{\infty} [\phi_i^Q \mathrm{B}_{t-i}^Q + \phi_i^T \mathrm{B}_{t-i}^T + \phi_i^{Q\&T} \mathrm{B}_{t-i}^Q \mathrm{B}_{t-i}^T + \beta_i] q_{t-i} + \varepsilon_t' \tag{4.6}$$

where the innovation ε_t, market maker's information, in Equation (4.2) is now separated into

$$\varepsilon_t = \underbrace{\sum_{i=0}^{\infty} [\phi_i^Q \mathrm{B}_{t-i}^Q + \phi_i^T \mathrm{B}_{t-i}^T + \phi_i^{Q\&T} \mathrm{B}_{t-i}^Q \mathrm{B}_{t-i}^T] q_{t-i}}_{\text{Information impounded into prices during bursts in quotes or trades}} + \underbrace{\varepsilon_t'}_{\text{Market maker's information after controlling for bursts}} \tag{4.7}$$

[31] In the high-frequency setting, we may interpret *information* as any knowledge regarding the likelihood of price movement in a time horizon relevant to a high-frequency trader, e.g. a few seconds. This certainly differs from information with price impact measured in days, e.g. information possessed by some institutional traders (Schedule 13D filers) discussed in Collin-Dufresne and Fos (2015).

[32] Throughout the paper we impose minimal econometric assumptions on the models considered. The data generating process $\{r_t, q_t, B_t^Q, \mathrm{B}_t^T, \varepsilon_t, \nu_t\}$ necessarily features heteroskedasticity and serial correlation. For consistent estimation and robust inference, we assume appropriate mixing conditions hold. Informally, mixing conditions describe asymptotic independence where events sufficiently far apart in time are close to being independent. See White (2014), pp. 43-45. This is certainly reasonable in the high-frequency microstructure setting. This is in contrast with Hasbrouck (1991) and later empirical investigations (e.g. Hendershott et al. (2011)) that impose covariance-stationarity.

Consistency with Equation (4.2) therefore requires that bursts in quotes and bursts in trades are uncorrelated with $\{r_{t-1}, r_{t-2}, \cdots\}$ and $\{q_t, q_{t-1}, q_{t-2}, \cdots\}$. This assumption rests on the transient nature, both in duration and effect on the market, of bursts in quotes and trades. The results in Section 4.5 show that these conditions are met[33]: the average duration of burst in quotes is 1.09 seconds and for trades 1.11 seconds. Quoted spread, effective spread, return volatility, and average order size all show concurrent spikes of the same duration. Bursts are therefore surprise events unanticipated by the rest of the market and exogenous to normal price updates and order flow.[34]

The order flow time series remains the same as Equation (4.3). Our model subsumes the Hasbrouck model as a special case when there is no burst activity. In particular, ε'_t, the market maker's information after controlling for bursts, and ν_t are contemporaneously uncorrelated and exogenous. As expected, our model selection procedure for choosing the number of lags confirms that burst indicators are not significant for lags greater than zero. We compare the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for models in Equation (4.2) and Equation (4.3) with up to 30 lags of past returns and order flow. Figure 4.7 reports AIC and BIC for one trading day of AA stock. Similar results are obtained for the rest of the stocks and days. Based on the values of AIC and BIC, for all stocks we choose the ten lags for r_t and q_t and zero lag for bursts in quotes and trades. We therefore arrive at the model

$$r_t = \sum_{i=1}^{10} \alpha_i r_{t-i} + \sum_{i=1}^{10} \beta_i q_{t-i} + \beta_0 q_t + \phi^Q \mathrm{B}_t^Q q_t + \phi^T \mathrm{B}_t^T q_t + \phi^{Q\&T} \mathrm{B}_t^Q \mathrm{B}_t^T q_t + \varepsilon'_t, \tag{4.8}$$

$$q_t = \sum_{i=1}^{10} \gamma_i r_{t-i} + \sum_{i=1}^{10} \delta_i q_{t-i} + \nu_t \tag{4.9}$$

Estimation results for this model are reported in Table 4.3.[35] The coefficient of bursts in quotes, ϕ^Q, has an average value of *0.009*, and is positive and significant at the 5% level during 81.67% of firm-days. The average impact of bursts in quotes is 50% larger than the corresponding impact of bursts in trades, which averages *0.006* and is significant during 91.67% of firm-days. Aggressive liquidity provision has an impact on prices, which is in its magnitude even larger than the impact of bursts in trades. The

[33] Section 4.7 provides further confirmation on the trasient nature of bursts.

[34] The behavior of the empirical adverse selection proxy in Section 4.7 also confirms that the shifts in the informational structure of the market associated to bursts in quotes are short-lived with no lag effects.

[35] Estimation results for the return Equation (4.2) and order flow Equation (4.3) are reported in Table 4.6 and Table 4.7, respectively, in the Appendix.

coefficient of concurrent bursts in quotes and trades, $\phi^{Q\&T}$, is most of the time not significant and its value is much smaller, *-0.002*. This suggests that the impact of bursts can be decomposed into the impact of quotes and trades without introducing the interaction term. The immediate impact of trades during one-second intervals, β_0, is *0.002* on average. Taken together with the impact of bursts in quotes, the immediate impact of trades increases to *0.011* during bursts in quotes, which is about five times higher than during no-burst periods.

4.6.2. Breakdown of Information Dichotomy During Bursts

The additional price impact during bursts in quotes observed in the previous subsection suggests a temporary structural break in the informational dichotomy between market makers and market takers. At the high-frequency resolution we consider, some information can be accessed by both market makers and market takers. There may be temporary contemporaneous correlation between market makers' and market takers' information, which in turn leads to a temporary change in the severity of adverse selection faced by the market maker, resulting in magnified price impact of trades we observe during bursts in quotes. We now test this hypothesis.

Econometrically, our hypothesis is that $\{\varepsilon_t, \nu_t\}$ are contemporaneously correlated during, and only during, bursts in quotes. To test for correlation during bursts in quotes, we take the residuals $\{\hat{\varepsilon}_t, \hat{\nu}_t\}$ from estimating Equation (4.2) and Equation (4.3). We retain those residuals

$$\{\hat{\varepsilon}_t \mathbb{1}_{\{\mathrm{B}_t^T=0\,\&\,\mathrm{B}_t^Q=1\}}, \hat{\nu}_t \mathbb{1}_{\{\mathrm{B}_t^T=0\,\&\,\mathrm{B}_t^Q=1\}}\} \tag{4.10}$$

from burst-in-quote periods and discard those from periods of bursts in trades.[36]

To test correlations during non-burst-in-quote periods, the non-burst-in-quote residuals $\hat{\varepsilon}_t \mathbb{1}_{\{\mathrm{B}_t^T=0\,\&\,\mathrm{B}_t^Q=0\}}$ from Equation (4.2) are not viable as input. If we expect positive correlation of $\{\hat{\varepsilon}_t, \hat{\nu}_t\}$ during bursts in trades or quotes, by estimating Equation (4.2) we force negative correlation on $\{\hat{\varepsilon}_t, \hat{\nu}_t\}$ during non-burst periods, because the regression residuals have zero correlation on average. Hence we take the residuals $\hat{\varepsilon}'_t$ from Equation (4.8) during non-burst-in-quote periods (again discarding burst-in-trades periods) and test the correlation between $\{\hat{\varepsilon}'_t \mathbb{1}_{\{\mathrm{B}_t^T=0\,\&\,\mathrm{B}_t^Q=0\}}, \hat{\nu}_t \mathbb{1}_{\{\mathrm{B}_t^T=0\,\&\,\mathrm{B}_t^Q=0\}}\}$. Equivalently,

[36] During bursts in trades, ε_t necessarily contains market takers' private information. Therefore not discarding residuals $\hat{\varepsilon}_t \mathbb{1}_{\{\mathrm{B}_t^T=1\,\&\,\mathrm{B}_t^Q=1\}}$ from bursts-in-trade periods will contaminate our correlation test.

we can write the market maker's information ε_t in Equation (4.2) as

$$\varepsilon_t = \underbrace{\phi_i^T \mathrm{B}_t^T q_t + \phi_i^{Q\&T} \mathrm{B}_t^Q \mathrm{B}_t^T q_t}_{\text{Information impounded into prices during bursts in trades}} \tag{4.11}$$
$$+ \underbrace{\varepsilon_t^{\#}}_{\text{Market maker's information during bursts in quotes only}} + \underbrace{\varepsilon_t'}_{\text{Market maker's information during non-burst periods}} .$$

Discarding the first part of the residuals representing the information impounded into prices during bursts in trades leaves us with

$$\hat{\varepsilon}_t \mathbb{1}_{\{\mathrm{B}_t^T=0\}} = \hat{\varepsilon}_t^{\#} + \hat{\varepsilon}_t'. \tag{4.12}$$

where we use $\hat{\varepsilon}_t \mathbb{1}_{\{\mathrm{B}_t^T=0\,\&\,\mathrm{B}_t^Q=1\}}$ and $\hat{\varepsilon}_t' \mathbb{1}_{\{\mathrm{B}_t^T=0\,\&\,\mathrm{B}_t^Q=0\}}$ as empirical proxy for $\varepsilon_t^{\#}$ and ε_t' respectively. Nonparametric correlation tests performed on the two sets of residuals are reported in Table 4.4. The correlation of $\{\hat{\varepsilon}_t, \hat{\nu}_t\}$ is significant during bursts in quotes whereas the correlation outside of bursts in quotes are nonsignificant with large margins.[37] We also test our hypothesis assuming a linear model. Restricting to non-burst-in-trade periods and assuming a linear relationship between ε_t and ν_t,we estimate the regression

$$\hat{\nu}_t = \eta^Q \mathrm{B}_t^Q \hat{\varepsilon}_t + \eta^{\neg Q}(1 - \mathrm{B}_t^Q)\hat{\varepsilon}_t' + \nu_t'. \tag{4.13}$$

Our null hypotheses are that $\eta^Q \neq 0$ and $\eta^{\neg Q} = 0$.

The results from estimating this specification, with residuals as proxies, are reported in Table 4.4. We clearly fail to reject the null hypotheses. There is a strong relationship between residuals and thus market maker's and market taker's information that exists only during bursts in quotes. The coefficient is not distinguishable from zero otherwise. Our results confirm that, during very brief episodes of aggressive liquidity provision, the limit-order-market-order interaction forms a two-way information channel. Outside of burst periods limit orders passively absorb information from market orders. We next examine how this informational structural break differently impacts the two sides of the order book.

[37]Pearson r, Spearman ρ, and Kendall τ statistics are all significant with p-values below 10^{-3} and an average of approximately 0.4.

4.7. Asymmetric Adverse Selection

We have shown that during bursts in quotes, market maker's and market takers' otherwise orthogonal information sets become correlated and that the price impact of trades rises significantly. We now analyze the adverse selection faced by market makers from the bid and ask sides of the limit order book before, during, and after bursts in quotes. We compute the standard empirical measures used in examining adverse selection. The *realized spread* at time t, denoted by RS_t, is defined by

$$RS_t = 2q_t(P_t - M_{t+s}) \tag{4.14}$$

where P_t is the transaction price at time t, and M_{t+s} is the midpoint at time $t+s$ for some chosen time lag s. RS_t is the difference between the current deal price and the quoted midpoint at a future time. After a transaction at time t, price movements favorable to the market maker from t to $t+s$ results in positive RS_t. For example, after a buyer-initiated trade ($q_t = 1$), if $P_t > M_{t+s}$, the market maker can unwind his short position to make a profit of $P_t - M_{t+s}$ per share at $t+s$. Conversely, $P_t < M_{t+s}$ means that, marking to market, the market maker has suffered an adverse price move and must revise his estimate of the fundamental value accordingly. The *effective spread* at time t, as defined in Section 4.5, is

$$ES_t = 2q_t(P_t - M_t). \tag{4.15}$$

ES_t is a proxy for the market maker's revenue from supplying immediacy. The following limit order book statistic defined using realized spread and effective spread serves as a reasonable proxy for adverse selection of market makers (see Bessembinder (2003)):

$$AD_t = ES_t - RS_t = 2q_t(M_{t+s} - M_t) \tag{4.16}$$

i.e. AD_t is the revenue from supplying immediacy plus the loss due to adverse price moves, or put differently, the permanent adverse price move suffered by market makers supplying liquidity. The higher the loss from the realized spread is, the higher the AD_t is.

AD_t is measured at a chosen time interval s; we choose $s = 30$ seconds. We report our results at two different sampling frequencies, one second and five seconds. At one-second intervals, for each instance of burst in our data, we first find the realized spread time series $\{RS_t\}$ for 12 intervals before and after, then average across all bursts. A similar procedure applied to the five-second sampling frequency gives a window of 120 seconds for each statistic (12 intervals of five seconds before and after burst). To summarize our findings, for each company we convert the raw realized spreads into percentages of the mid-quote and then average across all companies.[38] The results for one- and five-second sampling frequencies are displayed in Figure 4.8 and Figure 4.9, respectively.[39] In both Figure 4.8 and Figure 4.9, the left and right columns show the results for bursts on the ask and bid side. Inspecting the middle panel of both figures shows that market makers on bid and ask sides experience the same realized spread, except during a few-second window around bursts. The realized spread spikes up for those market makers when the trade takes place on the same side as burst—indicating that they are anticipating order flow, consistent with our analysis in Section 4.6. On the other hand, market makers on the opposite side of a burst experience temporarily worse realized spread. The change in price impact occurs concurrently with an approximately 80% reduction in incoming market order size shown in Figure 4.6.

The same normalization as for the realized spread is applied to effective spread calculations. The result is shown in the bottom panels of Figure 4.8 and Figure 4.9. We observe that the effective spread increases during bursts at either side of the order flow. The amount of increase differs depending on whether the market maker is on the same side as the burst. The market makers on the opposite side of the burst charge higher cost for immediacy. This is consistent with our analysis above. Market makers charge higher transaction cost for order flow trading in the direction of the permanent price impact. For example, in the left columns of Figure 4.8 and Figure 4.9, during a burst in quotes on the ask side, the premium for immediacy charged against sell orders spikes up. We find it interesting that bursts in quotes, whose average duration is approximately one second, affect price movements five seconds after their occurrence.

Despite extracting a higher immediacy premium, market makers on the opposite side of a burst in quotes nevertheless suffer worse adverse selection relative to non-burst periods. The results of adverse selection proxy are given in the top panels of Figure 4.8 and Figure 4.9. During bursts there are clearly opposite shifts in adverse selection for bid- and ask-side market makers. The adverse selection proxy for the market makers on the opposite side as the burst shows an upward jump during burst, while a downward dip occurs for those on the same side. There is therefore asymmetry in adverse

[38]For example, the realized spread is converted to $\frac{RS_t}{M_t} \times 100\%$.

[39]In computing empirical proxies, we only consider bursts in quotes where bursts only occur on one side of the book. Table 4.5 provides details on the amount of data used in this procedure.

selection between the two sides of the limit order book during bursts.[40] While the market makers on the opposite side of the burst charge higher cost of immediacy for order flow trading in the direction of the permanent price impact, it only partially compensates for the negative realized spread resulting from these events.

The model Equation (4.8) can also be extended to this higher resolution setting, by distinguishing bursts in quotes on the bid and ask sides. The directional specification is[41]

$$r_t = \sum_{i=1}^{10} \alpha_i r_{t-i} + \sum_{i=1}^{10} \beta_i q_{t-i} + \beta_0 q_t + \phi^{Q,same} \mathrm{B}_t^{Q,same} q_t + \phi^{Q,opp} \mathrm{B}_t^{Q,opp} q_t + \phi^T \mathrm{B}_t^T q_t + \varepsilon_t', \tag{4.17}$$

where r_t is again the percentage mid-quote return during period t. The indicator $\mathrm{B}_t^{Q,same}$ is 1 if q_t hits the side of the book where there is a burst—e.g. a buyer-initiated order during a burst on the ask side—and zero otherwise. Similarly, the indicator $\mathrm{B}_t^{Q,opp}$ is 1 if q_t hits the side of the book opposite a burst. Results are reported in Table 4.5. The coefficient $\phi^{Q,same}$ is negative and significant at the 5% confidence level for 67.33% of firm-days. The coefficient $\phi^{Q,opp}$ is positive and significant at the 5% confidence level for 82.67% of firm-days. The results confirm the asymmetry in adverse selection during bursts. Our previous findings from Section 4.6 that market makers' and market takers' information become correlated during bursts in quotes can be made more precise. The brief information advantage is only enjoyed by those market makers who precipitate the burst. The orders on the opposite side of the burst in quotes have elevated permanent price impact. On the other hand, for order flow hitting the same side of the book as a burst, the price impact is negative and possibly profitable for the market maker.

4.8. Conclusions

In this paper, we analyze periods of extreme behavior of the quotation process in the U.S. equity market. We show that these very brief periods are associated with sharp spikes in return volatility, effective spread, and quoted spread, and with a significant decrease in order size. Despite only occurring in sporadic seconds-long episodes exogenous to normal order flow, bursts in quotes impart permanent price impact. Moreover, the price impact of bursts in quotes is economically more significant than an

[40] As a byproduct, these results also serve as further confirmation that bursts are exogenous shocks largely unanticipated by the market in general, which was assumed in Equation (4.8).

[41] All pairwise interaction terms between dummies are included in the regression but suppressed in the equation for notional compactness.

analogous impact of bursts in trades. Further analysis shows that bursts in quotes are indications of short-lived structural breaks in information asymmetry between market maker and market taker, a departure from the standard microstructure paradigm.

Our analysis calls attention to a phenomenon thus far unexamined by the microstructure literature—price impact due to market maker action. A sub-population of market makers—those that initiate bursts—enjoy a temporary abatement of adverse selection during bursts. The rest of market makers suffers more severe adverse selection relative to non-burst periods. One may put forth a number of conjectures regarding the reason behind this fleeting merge of information from two sides of the market. High-frequency market makers may use their speed and information processing advantage to minimize adverse selection by predicting the direction of price moves. For example, during public news announcements, high-frequency traders may be using text-mining technology to anticipate market moves (see e.g. Foucault et al. (2015) and Roşu (2014)). Or market takers may be using limit orders as part of their optimal execution strategy. Another possibility is that bursts-in-quote may be a tactic designed to induce response from other market makers, in which case the resulting asymmetric adverse selection would indicate this is a zero-sum game among market makers.

Table 4.1
Data Summary Statistics

This table reports daily medians of 15 highly liquid stocks from October 10, 2011 to February 7, 2012 for 82 trading days. The data is obtained from the Trades and Quotes (TAQ) database. The number of trades is obtained as the number of transactions from the TAQ. The number of quotes represents changes in the National Best Bid and Offer (NBBO) either in price or size. Price is the volume weighted average price of eligible transactions. Spread is given by the difference between ask and bid price. Size is measured in number of lots (100 shares). Ticker symbols and company names are provided in the first two columns.

Stock	Company Name	# of trades	# of quotes	Price	Size in lots	Spread in ¢
AA	Alcoa	53,207	1,436,262	9.87	6,003	1.00
BAC	Bank of America	136,476	1,767,462	6.25	44,121	1.00
GE	General Electric	81,607	2,008,788	17.21	9,128	1.00
IBM	International Business Machines	17,838	754,269	186.05	840	3.04
JNJ	Johnson & Johnson	29,928	1,487,517	64.48	1,582	1.63
JPM	JPMorgan Chase	90,824	3,501,468	33.79	5,364	1.35
MRK	Merck & Co	39,242	1,707,993	35.99	2,520	1.13
PFE	Pfizer	68,399	1,802,107	20.44	6,646	1.00
PG	Procter & Gamble	27,544	1,404,016	64.62	1,486	1.71
PM	Philip Morris International	19,413	952,338	73.60	1,092	2.22
T	AT&T	51,391	1,699,231	29.32	3,926	1.01
VZ	Verizon Communications	34,330	1,444,768	37.82	2,279	1.11
WFC	Wells Fargo	82,357	3,000,480	27.04	5,046	1.20
WMT	Wal-Mart Stores	30,718	1,409,885	58.49	1,683	1.33
XOM	Exxon Mobil	57,168	4,734,158	81.38	2,512	1.90

Table 4.2
Summary Statistics for Bursts

Median daily statistics of bursts in quotes for the 20 most active stocks on NASDAQ and NYSE from March 9, 2015 to April 8, 2015 using 100-millisecond (ms) intervals and the 15 most active NYSE stocks from October 10, 2011 to February 7, 2012 using one-second and five-second intervals. We compute the average length and duration between bursts for every day and report the medians across days. The threshold for bursts is 99.9%.

Interval length	Median # of bursts	Average length	Average duration between bursts	% Volume during bursts	% of time in a day
100 milliseconds	5,692.1	0.12s	3.25s	54.7%	2.9%
One second	419.3	1.09s	42.6s	6.4%	2.3%
Five seconds	24.0	5.27s	1536.7s	1.2%	0.6%

Table 4.3

Estimated Coefficients for the Price Impact Equation with Bursts

This table shows estimated coefficients for the basic VAR specification with bursts in quotes, bursts in trades, and bursts in both quotes and trades. Heteroskedasticity and autocorrelation adjusted p-values are displayed below each coefficient. r_t is the percentage mid-quote return based on the average mid-quote immediately before t and immediately after t. q_t is the signed trade direction and B_t^Q is a dummy variable indicating the presence of bursts in quotes in the interval t. Coefficient ϕ^Q represents the price impact attributed to the burst in quotes, B_t^Q. Coefficient ϕ^T represents the price impact attributed to the burst in trades, B_t^T. Coefficient $\phi^{Q\&T}$ represents the price impact attributed to the burst in quotes with concurrent burst in trades, $\mathrm{B}_t^Q\mathrm{B}_t^T$. We display only the leading terms of the lagged returns. The estimation is performed for each company for every day. We report the median value of coefficients across stocks with the associated p-value and the percentage of companies for which a specific coefficient is significant across days. We also display the median number of each type of burst event per day. The analysis in this table is based on one-second intervals and a 99.9% threshold for bursts. The sample period is October 10, 2011 to February 7, 2012.

$$r_t = \sum_{i=1}^{10} \alpha_i r_{t-i} + \sum_{i=1}^{10} \beta_i q_{t-i} + \beta_0 q_t + \phi^Q \mathrm{B}_t^Q q_t + \phi^T \mathrm{B}_t^T q_t + \phi^{Q\&T} \mathrm{B}_t^Q \mathrm{B}_t^T q_t + \varepsilon_t'.$$

	α_1	β_0	ϕ^Q	ϕ^T	$\phi^{Q\&T}$	R^2-Adj with bursts	R^2-Adj w/o bursts
AA	0.925	0.002	0.046	0.007	-0.012	0.585	0.489
	0.000	0.000	0.000	0.000	0.030		
BAC	0.988	0.001	0.058	0.007	-0.009	0.593	0.510
	0.000	0.000	0.000	0.001	0.139		
GE	0.917	0.001	0.028	0.007	-0.007	0.567	0.484
	0.000	0.000	0.000	0.000	0.021		
IBM	0.247	0.003	0.006	0.005	-0.001	0.416	0.416
	0.000	0.000	0.022	0.013	0.281		
JNJ	0.445	0.002	0.004	0.004	-0.001	0.411	0.397
	0.000	0.000	0.009	0.000	0.213		
JPM	0.875	0.006	0.009	0.011	-0.002	0.502	0.482
	0.000	0.000	0.005	0.000	0.254		
MRK	0.851	0.002	0.012	0.007	-0.004	0.518	0.465
	0.000	0.000	0.000	0.000	0.045		
PFE	0.827	0.000	0.023	0.005	-0.001	0.541	0.461
	0.000	0.000	0.000	0.000	0.125		
PG	0.619	0.002	0.005	0.005	-0.001	0.403	0.381
	0.000	0.000	0.000	0.000	0.210		
PM	0.336	0.002	0.004	0.006	0.000	0.407	0.407
	0.000	0.043	0.113	0.009	0.192		
T	0.742	0.001	0.015	0.006	-0.006	0.484	0.430
	0.000	0.000	0.000	0.000	0.027		
VZ	0.846	0.001	0.009	0.007	-0.003	0.505	0.437
	0.000	0.000	0.000	0.000	0.076		
WFC	0.922	0.005	0.013	0.013	-0.008	0.551	0.509
	0.000	0.000	0.000	0.000	0.060		
WMT	0.705	0.002	0.007	0.005	-0.002	0.462	0.430
	0.000	0.000	0.000	0.000	0.082		
XOM	0.463	0.002	0.004	0.006	0.002	0.459	0.441
	0.000	0.000	0.138	0.000	0.256		
Median coefficient	0.827	0.002	0.009	0.006	-0.002	0.502	0.441
Median p-value	0.000	0.000	0.000	0.000	0.125		
Median daily # of events			419.3	706.1	318.0		
% of significant firm-days at 5%	99.67%	93.00%	81.67%	91.67%	30.00%		

Table 4.4

Estimates of Information Relationship During Bursts in Quotes

This table shows Pearson's correlations ρ and regression estimates of the relation between market takers' and market makers' information, ε_t and ν_t. ρ^Q refers to correlation between $\hat{\varepsilon}_t$ and $\hat{\nu}_t$ only during bursts in quotes, and $\rho^{\neg Q}$ to correlation outside of bursts in quotes or trades periods based on residuals from Equation (4.8) controlling for the effect of bursts. Using residuals from Equation (4.2) and controlling for bursts in trades, the correlation between $\hat{\varepsilon}_t$ and $\hat{\nu}_t$ outside of bursts in trades is ρ^{All}. η^Q and $\eta^{\neg Q}$ denote corresponding OLS estimates from the equation

$$\hat{\nu}_t = \eta^Q \mathrm{B}_t^Q \hat{\varepsilon}_t + \eta^{\neg Q}(1 - \mathrm{B}_t^Q)\hat{\varepsilon}'_t + \nu'_t.$$

The estimation is performed for each company for every day. We report the median value of coefficients across stocks with the associated p-value and the percentage of companies for which a specific coefficient is significant across days. We also display the median daily number of observations we use for computing a correlation coefficient and this corresponds to non-zero entries in an array used in the regression. The analysis in this table is based on one-second intervals and a 99.9% threshold for bursts. The sample period is October 10, 2011 to February 7, 2012.

	ρ^Q	$\rho^{\neg Q}$	η^Q	$\eta^{\neg Q}$	ρ^{All}
AA	0.487	-0.002	7.779	-0.130	-0.002
	0.000	0.724	0.000	0.702	0.726
BAC	0.431	0.007	4.230	0.508	0.005
	0.000	0.292	0.000	0.411	0.427
GE	0.514	-0.002	12.735	-0.237	-0.002
	0.000	0.729	0.000	0.696	0.720
IBM	0.193	0.000	7.789	0.000	0.000
	0.000	0.980	0.001	0.984	0.982
JNJ	0.323	0.000	20.476	-0.021	0.000
	0.000	0.947	0.000	0.936	0.946
JPM	0.413	0.000	13.601	-0.037	0.000
	0.000	0.905	0.000	0.923	0.913
MRK	0.516	-0.002	22.172	-0.247	-0.002
	0.000	0.744	0.000	0.716	0.765
PFE	0.478	0.000	13.568	-0.087	-0.001
	0.000	0.779	0.000	0.784	0.828
PG	0.339	0.000	22.921	-0.049	0.000
	0.000	0.922	0.000	0.900	0.935
PM	0.022	0.000	0.526	-0.001	0.000
	0.000	0.970	0.110	0.971	0.970
T	0.478	-0.001	21.501	-0.202	-0.001
	0.000	0.782	0.000	0.788	0.809
VZ	0.417	-0.003	21.597	-0.332	-0.002
	0.000	0.691	0.000	0.652	0.716
WFC	0.493	-0.001	15.465	-0.060	0.000
	0.000	0.862	0.000	0.871	0.884
WMT	0.410	0.000	26.169	-0.017	0.000
	0.000	0.922	0.000	0.937	0.921
XOM	0.206	0.000	15.633	-0.046	0.000
	0.000	0.934	0.128	0.935	0.934
Median coefficient	0.417	0.000	15.465	-0.049	0.000
Median p-value	0.000	0.862	0.000	0.871	0.884
Median daily # of events	101.3	20,789	101.3	20,789	20,890
% of significant firm-days at 5%	100%	0%	100%	0%	0%

Table 4.5

Estimated Coefficients for the Directed Price Impact Equation in the VAR

This table reports estimated coefficients for the basic VAR specification with bursts in quotes in the direction of the trade, in the opposite direction of the trade, and bursts in trades. Heteroskedasticity and autocorrelation adjusted p-values are displayed below each coefficient. r_t is the percentage mid-quote return based on the average mid-quote immediately before t and immediately after t. q_t is the signed trade direction and B_t is a dummy variable indicating the presence of bursts in the interval t. Coefficient $\phi^{Q,same}$ represents the price impact attributed to the burst on the same side of the book as the sign of the order flow and no burst on the opposite side of the book, $\mathrm{B}_t^{Q,same}$. Hence, $\mathrm{B}_t^{Q,same}$ is 1 if q_t hits the side of the book where there is a burst, e.g. a buyer-initiated order during a burst on the ask side and zero otherwise. Coefficient $\phi^{Q,opp}$ represents the price impact attributed to the burst in quotes on the opposite side of the book. Hence, $\mathrm{B}_t^{Q,opp}$ is 1 if q_t hits the side of the book opposite a burst. Coefficient ϕ^T represents the price impact attributed to the burst in trades, B_t^T. We display only the leading terms for r_t and q_t. The estimation is performed for each company for every day. We report the median value of coefficients across stocks with the associated p-value and the percentage of companies for which a specific coefficient is significant across days. We also display the median number of each type of burst event per day. The analysis in this table is based on one-second intervals and a 99.9% threshold for bursts. The sample period is October 10, 2011 to February 7, 2012.

$$r_t = \sum_{i=1}^{10} \alpha_i r_{t-i} + \sum_{i=1}^{10} \beta_i q_{t-i} + \beta_0 q_t + \phi^{Q,same}\mathrm{B}_t^{Q,same} q_t + \phi^{Q,opp}\mathrm{B}_t^{Q,opp} q_t + \phi^T \mathrm{B}_t^T q_t + \varepsilon_t'.$$

	α_1	β_0	$\phi^{Q,same}$	$\phi^{Q,opp}$	ϕ^T	R^2-Adj with bursts	R^2-Adj w/o bursts
AA	0.929	0.002	-0.030	0.041	0.031	0.545	0.489
	0.000	0.000	0.000	0.000	0.000		
BAC	0.988	0.001	-0.024	0.040	0.043	0.558	0.510
	0.000	0.000	0.002	0.003	0.000		
GE	0.913	0.001	-0.025	0.030	0.018	0.534	0.484
	0.000	0.000	0.000	0.000	0.000		
IBM	0.247	0.003	0.003	0.007	0.007	0.416	0.416
	0.000	0.000	0.178	0.046	0.000		
JNJ	0.445	0.002	-0.006	0.008	0.005	0.413	0.397
	0.000	0.000	0.007	0.001	0.000		
JPM	0.875	0.006	-0.008	0.010	0.013	0.504	0.482
	0.000	0.000	0.010	0.000	0.000		
MRK	0.848	0.002	-0.014	0.011	0.009	0.507	0.465
	0.000	0.000	0.000	0.000	0.000		
PFE	0.826	0.001	-0.019	0.024	0.017	0.519	0.461
	0.000	0.000	0.000	0.000	0.000		
PG	0.619	0.002	-0.003	0.005	0.006	0.397	0.381
	0.000	0.000	0.125	0.000	0.000		
PM	0.336	0.002	0.000	0.001	0.006	0.408	0.407
	0.000	0.035	0.205	0.100	0.000		
T	0.742	0.001	-0.013	0.016	0.011	0.472	0.430
	0.000	0.000	0.000	0.000	0.000		
VZ	0.845	0.001	-0.010	0.010	0.009	0.486	0.437
	0.000	0.000	0.000	0.000	0.000		
WFC	0.920	0.005	-0.013	0.016	0.014	0.547	0.509
	0.000	0.000	0.001	0.000	0.000		
WMT	0.704	0.002	-0.008	0.009	0.006	0.455	0.430
	0.000	0.000	0.003	0.000	0.000		
XOM	0.446	0.002	0.002	0.002	0.006	0.455	0.441
	0.000	0.000	0.111	0.138	0.000		
Median coefficient	0.826	0.002	-0.010	0.010	0.009	0.486	0.441
Median p-value	0.000	0.000	0.002	0.000	0.000		
Median daily number of events			24.8	60.7	318		
% of significant firm-days at 5%	99.0%	93.33%	67.33%	82.67%	96.0%		

Table 4.6

Estimated Coefficients for the Return Equation in the VAR

This table reports estimated coefficients for the basic VAR specification without bursts in quotes. Heteroskedasticity and autocorrelation adjusted p-values are displayed below each coefficient. r_t is the percentage mid-quote return based on the average mid-quote immediately before t and immediately after t. q_t is the signed trade direction. The estimation is performed for each company for every day. We report the median value of coefficients across stocks with the associated p-value and the percentage of companies for which a specific coefficient is significant across days. The analysis in this table is based on one-second intervals. The sample period is October 10, 2011 to February 7, 2012.

$$r_t = \sum_{i=1}^{10} \alpha_i r_{t-i} + \sum_{i=1}^{10} \beta_i q_{t-i} + \beta_0 q_t + \varepsilon_t.$$

	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	α_{10}	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	R^2-Adj
AA	0.925	-0.853	0.744	-0.661	0.563	-0.474	0.374	-0.279	0.179	-0.091	0.001	0.002	0.000	0.001	0.000	0.001	0.000	0.001	-0.001	0.001	0.000	0.489
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.238	0.000	0.062	0.000	0.033	0.000	0.004	0.000	0.006	
BAC	0.989	-0.906	0.794	-0.695	0.589	-0.488	0.389	-0.300	0.190	-0.095	0.001	0.001	0.000	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.510
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.027	0.000	0.178	0.000	0.194	0.007	0.042	0.006	0.065	
GE	0.915	-0.844	0.738	-0.659	0.551	-0.460	0.368	-0.275	0.183	-0.092	0.001	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.484
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.168	0.000	0.085	0.000	0.051	0.001	0.011	0.001	0.012	
IBM	0.247	-0.829	0.208	-0.660	0.169	-0.493	0.110	-0.326	0.058	-0.159	0.002	0.003	0.003	0.004	0.002	0.002	0.001	0.002	0.001	0.001	0.000	0.416
	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.013	0.000	0.002	0.001	0.007	0.001	0.010	0.055	0.066	0.007	0.297	0.188	0.242	
JNJ	0.445	-0.771	0.368	-0.608	0.261	-0.445	0.216	-0.277	0.116	-0.119	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.397
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.031	0.001	0.017	0.001	0.079	0.016	0.181	0.051	0.142	
JPM	0.874	-0.790	0.707	-0.617	0.534	-0.438	0.352	-0.260	0.171	-0.084	0.005	0.002	0.000	0.001	0.000	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.482
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.003	0.001	0.000	0.001	0.000	0.002	0.000	
MRK	0.848	-0.785	0.692	-0.602	0.515	-0.427	0.346	-0.259	0.170	-0.084	0.001	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.465
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.243	0.000	0.012	0.000	0.003	0.000	0.004	0.000	0.001	
PFE	0.828	-0.781	0.686	-0.606	0.515	-0.439	0.348	-0.258	0.171	-0.086	0.001	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.461
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.093	0.000	0.170	0.000	0.074	0.001	0.016	0.002	0.007	
PG	0.618	-0.718	0.533	-0.539	0.426	-0.375	0.298	-0.229	0.150	-0.087	0.002	0.002	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.381
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.059	0.000	0.102	0.011	0.070	0.008	0.035	
PM	0.336	-0.818	0.278	-0.656	0.212	-0.474	0.142	-0.310	0.077	-0.147	0.001	0.003	0.002	0.003	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.407
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.099	0.004	0.043	0.005	0.018	0.020	0.055	0.029	0.134	0.109	0.139	
T	0.742	-0.722	0.623	-0.552	0.471	-0.396	0.316	-0.237	0.159	-0.088	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.430
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.050	0.000	0.166	0.001	0.108	0.002	0.080	0.002	0.084	
VZ	0.847	-0.783	0.678	-0.597	0.506	-0.426	0.339	-0.256	0.170	-0.087	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.437
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.161	0.000	0.073	0.000	0.012	0.002	0.010	0.009	0.004	
WFC	0.920	-0.820	0.739	-0.634	0.552	-0.456	0.363	-0.270	0.180	-0.087	0.005	0.002	0.000	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.509
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.098	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
WMT	0.704	-0.735	0.610	-0.548	0.454	-0.385	0.311	-0.231	0.157	-0.080	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.430
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.026	0.000	0.092	0.001	0.042	0.054	0.030	0.031	0.022	
XOM	0.541	-0.789	0.430	-0.621	0.321	-0.435	0.251	-0.262	0.145	-0.096	0.002	0.002	0.001	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.441
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.004	0.001	0.031	0.010	0.085	0.005	0.090	
Median coeff	0.828	-0.789	0.678	-0.617	0.506	-0.439	0.339	-0.262	0.170	-0.088	0.001	0.002	0.000	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.441
Median p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.050	0.000	0.059	0.000	0.051	0.002	0.030	0.005	0.022	

Table 4.7

Estimated Coefficients for the Order Flow Equation in the VAR

This tables shows estimated coefficients for the basic VAR specification without bursts in quotes. Heteroskedasticity and autocorrelation adjusted p-values are displayed below each coefficient. r_t is the percentage mid-quote return based on the average mid-quote immediately before t and immediately after t. q_t is the signed trade direction. The estimation is performed for each company for every day. We report the median value of coefficients across stocks with the associated p-value and the percentage of companies for which a specific coefficient is significant across days. The analysis in this table is based on one-second intervals. The sample period is October 10, 2011 to February 7, 2012.

$$q_t = \sum_{i=1}^{10} \gamma_i r_{t-i} + \sum_{i=1}^{10} \delta_i q_{t-i} + \nu_t.$$

	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9	γ_{10}	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6	δ_7	δ_8	δ_9	δ_{10}	R^2-Adj
AA	-3.320	0.577	-1.896	1.006	-1.293	0.773	-0.803	0.491	-0.404	0.327	0.096	0.065	0.041	0.038	0.032	0.033	0.023	0.026	0.026	0.026	0.036
	0.000	0.162	0.000	0.030	0.019	0.200	0.147	0.226	0.299	0.244	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	
BAC	-3.999	0.862	-2.392	1.287	-1.756	1.030	-1.270	0.619	-0.687	0.295	0.119	0.084	0.064	0.053	0.046	0.043	0.038	0.037	0.034	0.034	0.071
	0.000	0.049	0.000	0.034	0.004	0.078	0.024	0.242	0.132	0.295	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
GE	-6.843	1.575	-3.995	2.498	-2.642	2.127	-1.788	1.244	-1.269	0.926	0.101	0.065	0.045	0.037	0.033	0.030	0.022	0.023	0.023	0.025	0.041
	0.000	0.024	0.000	0.005	0.005	0.024	0.024	0.097	0.095	0.069	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.001	0.001	0.000	
IBM	-0.071	-0.006	-0.067	-0.004	-0.052	-0.012	-0.038	-0.009	-0.025	-0.015	0.054	0.023	0.014	0.011	0.012	0.008	0.006	0.008	0.008	0.011	0.006
	0.048	0.414	0.145	0.395	0.258	0.471	0.343	0.479	0.374	0.459	0.000	0.001	0.049	0.101	0.089	0.219	0.268	0.211	0.233	0.119	
JNJ	-2.733	0.047	-1.762	0.227	-1.153	0.206	-0.441	0.110	-0.232	-0.015	0.069	0.033	0.019	0.018	0.017	0.014	0.011	0.010	0.011	0.014	0.011
	0.000	0.176	0.001	0.163	0.036	0.211	0.176	0.381	0.280	0.322	0.000	0.000	0.005	0.009	0.014	0.039	0.103	0.121	0.108	0.047	
JPM	-4.274	1.308	-2.319	1.512	-1.676	1.114	-1.077	0.831	-0.491	0.267	0.105	0.050	0.028	0.024	0.019	0.017	0.012	0.015	0.011	0.016	0.024
	0.000	0.007	0.000	0.011	0.007	0.067	0.049	0.124	0.251	0.277	0.000	0.000	0.000	0.001	0.008	0.014	0.081	0.037	0.115	0.020	
MRK	-6.778	2.090	-3.996	2.476	-2.735	2.129	-1.849	1.320	-0.929	0.669	0.092	0.052	0.033	0.028	0.026	0.024	0.014	0.017	0.016	0.019	0.026
	0.000	0.007	0.000	0.004	0.003	0.027	0.042	0.104	0.190	0.197	0.000	0.000	0.000	0.000	0.000	0.001	0.044	0.012	0.019	0.006	
PFE	-6.710	1.279	-3.666	2.017	-2.318	1.775	-1.592	1.294	-0.912	0.945	0.102	0.072	0.048	0.045	0.039	0.036	0.029	0.031	0.028	0.033	0.049
	0.000	0.063	0.000	0.020	0.015	0.042	0.050	0.116	0.152	0.049	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
PG	-4.726	1.038	-2.952	1.420	-1.988	1.190	-1.023	0.514	-0.570	0.264	0.070	0.036	0.021	0.023	0.017	0.016	0.010	0.012	0.011	0.013	0.013
	0.000	0.094	0.000	0.075	0.015	0.117	0.102	0.315	0.202	0.357	0.000	0.000	0.002	0.001	0.013	0.018	0.165	0.098	0.097	0.067	
PM	-0.160	-0.046	-0.160	-0.045	-0.139	-0.029	-0.095	-0.017	-0.044	-0.018	0.054	0.017	0.010	0.010	0.013	0.007	0.007	0.008	0.006	0.011	0.006
	0.000	0.126	0.003	0.249	0.044	0.271	0.154	0.378	0.310	0.440	0.000	0.012	0.126	0.143	0.052	0.269	0.277	0.216	0.179	0.115	
T	-6.954	1.504	-4.118	2.332	-2.917	1.888	-1.747	1.311	-1.020	0.849	0.081	0.055	0.039	0.032	0.032	0.027	0.022	0.023	0.023	0.027	0.028
	0.000	0.070	0.000	0.033	0.007	0.067	0.070	0.166	0.266	0.175	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.000	
VZ	-7.679	2.228	-4.398	2.649	-3.118	2.328	-2.298	1.415	-1.147	0.833	0.075	0.045	0.029	0.026	0.025	0.020	0.016	0.019	0.017	0.022	0.021
	0.000	0.007	0.000	0.006	0.002	0.035	0.029	0.090	0.134	0.185	0.000	0.000	0.000	0.000	0.000	0.003	0.022	0.008	0.014	0.002	
WFC	-4.495	1.116	-2.228	1.642	-1.706	1.289	-1.166	0.806	-0.560	0.317	0.107	0.059	0.036	0.032	0.029	0.024	0.018	0.019	0.017	0.021	0.03
	0.000	0.055	0.000	0.017	0.017	0.069	0.103	0.192	0.276	0.317	0.000	0.000	0.000	0.000	0.000	0.001	0.012	0.007	0.018	0.002	
WMT	-4.372	0.856	-2.439	1.157	-1.235	1.137	-0.972	0.626	-0.387	0.463	0.077	0.038	0.026	0.022	0.023	0.016	0.013	0.014	0.013	0.016	0.015
	0.000	0.136	0.002	0.186	0.069	0.198	0.272	0.283	0.336	0.281	0.000	0.000	0.000	0.001	0.001	0.022	0.054	0.042	0.065	0.021	
XOM	-4.964	1.137	-2.460	1.018	-1.707	0.771	-0.708	0.092	-0.322	-0.043	0.092	0.034	0.019	0.017	0.016	0.014	0.010	0.012	0.010	0.013	0.017
	0.000	0.034	0.000	0.102	0.010	0.178	0.063	0.300	0.273	0.329	0.000	0.000	0.005	0.016	0.024	0.048	0.169	0.088	0.149	0.054	
Median coeff	-4.495	1.116	-2.439	1.420	-1.707	1.137	-1.077	0.626	-0.560	0.317	0.092	0.050	0.029	0.026	0.025	0.020	0.014	0.017	0.016	0.019	0.024
Median p-value	0.000	0.063	0.000	0.033	0.015	0.078	0.070	0.226	0.266	0.281	0.000	0.000	0.000	0.000	0.000	0.003	0.044	0.012	0.019	0.006	

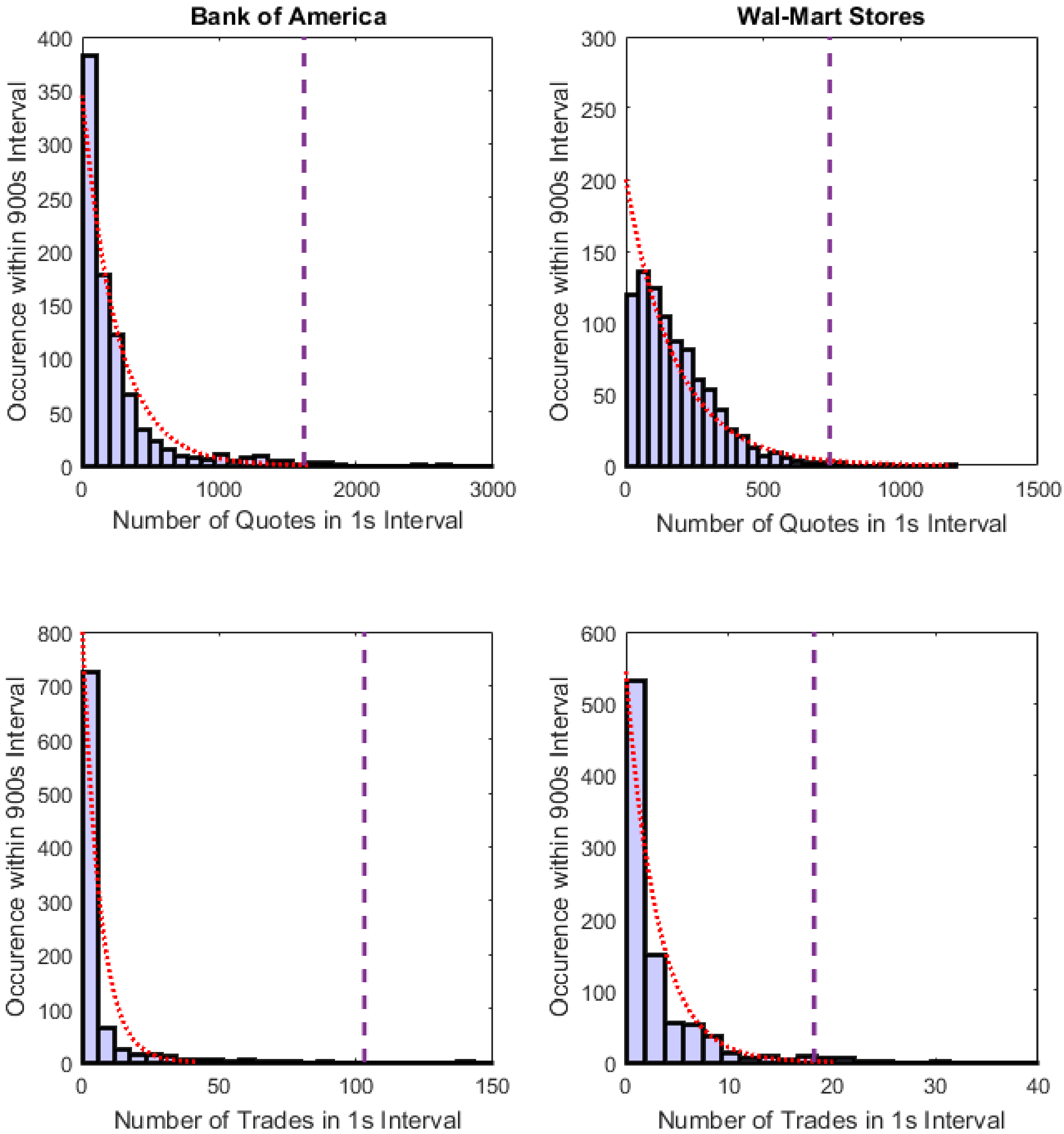

Figure 4.1: Burst threshold for the number of quotes and trades during 900 one-second intervals.

Thresholds for determining bursts during 900 intervals of length one-second are given by the dashed lines. The threshold for the 15-minute window is computed based on an exponential fit to the 25-minute interval including 5 minutes before and 5 minutes after the 15-minute interval. For every one-second interval once the corresponding series reaches the threshold, a burst for that interval is recorded. The top panels show the number of quotes and the bottom panels the number of trades. These graphs represent the data for Bank of America and Wal-Mart Stores (tickers BAC on the left and WMT on the right) between 10:00am and 10:15am on October 13, 2011. The fitted exponential distributions are shown in dotted lines. The estimated means of the exponential distributions based on the 15-minute intervals are 250.2 and 179.5 for the number of quotes for the respective stocks, and 6.3 and 3.0 for the number of trades.

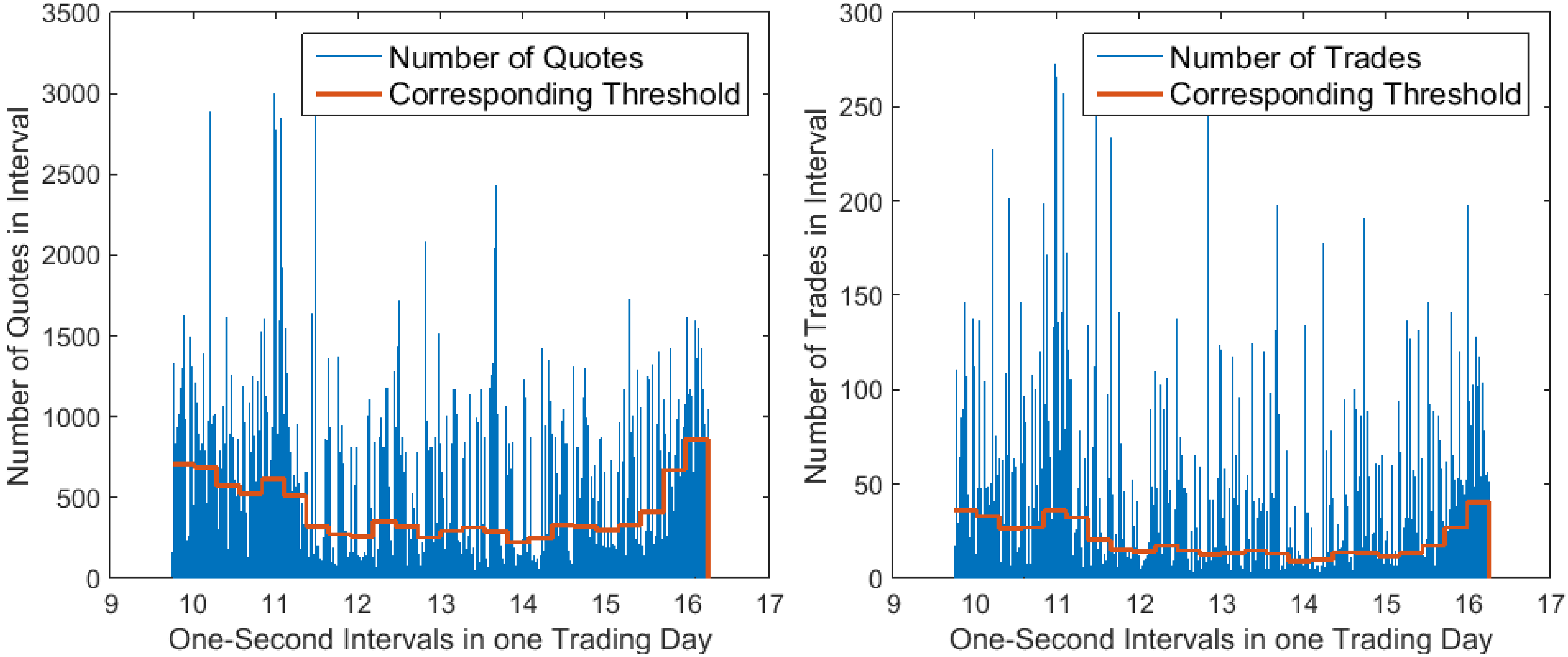

Figure 4.2: Burst threshold for the number of quotes and trades during one day.

Thresholds for determining bursts. For every one-second interval once the corresponding series reaches the threshold, a burst for that interval is recorded. The left panel shows the number of quotes and the right panel the number of trades. These graphs represent the data for Alcoa (ticker AA) between 9:45am and 15:45pm on October 10, 2011.

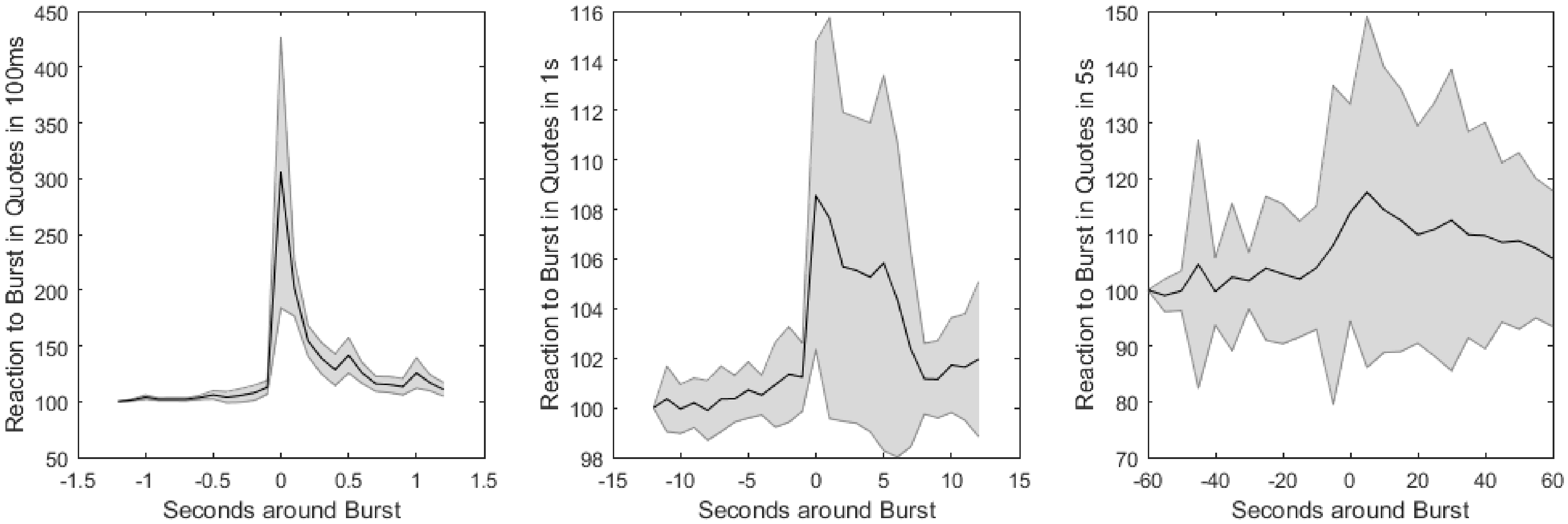

Figure 4.3: Quoted spread around bursts in quotes.

This figure shows the quoted spread in a 2.5-second (25-second, 2-minute) window around bursts in quotes on the left (middle, right). The black line represents the average across all 100-millisecond (1-second, 5-second) intervals before, during, and after bursts, as well as across all stocks in our sample, where the initial value was normalized to 100. The shaded area represents one standard deviation from the mean.

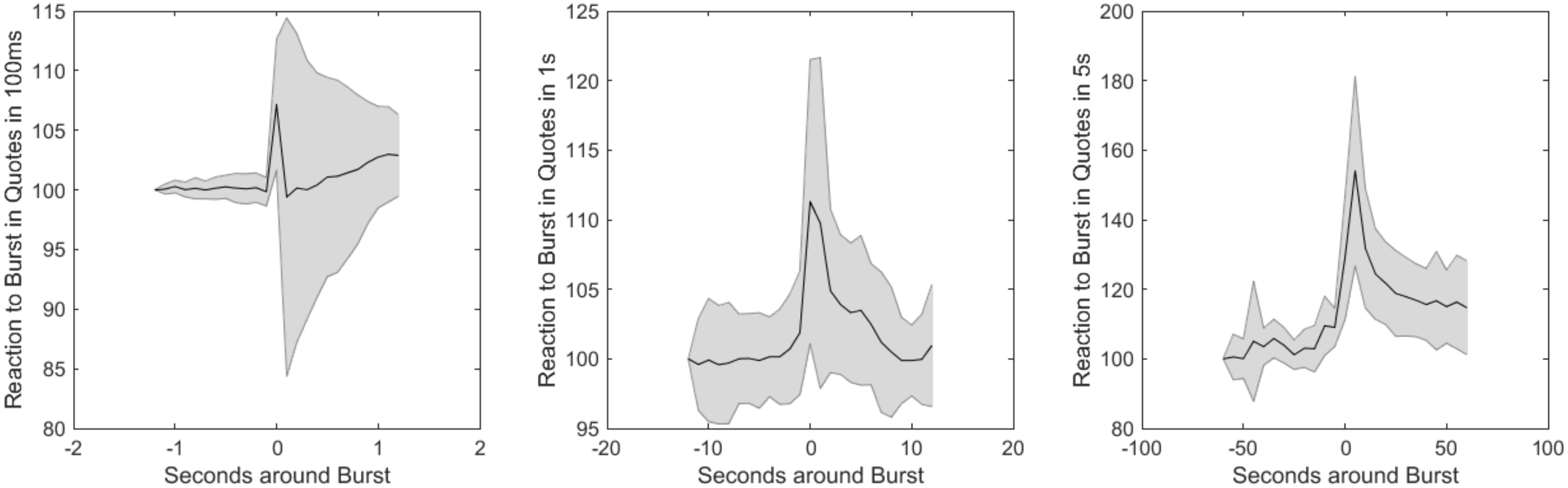

Figure 4.4: Effective spread around bursts in quotes.

This figure examines the effective spread in a 2.5-second (25-second, 2-minute) window around bursts in quotes on the left (middle, right). The black line represents the average across all 100-millisecond (1-second, 5-second) intervals before, during, and after bursts, as well as across all stocks in our sample, where the initial value was normalized to 100. The shaded area represents one standard deviation from the mean.

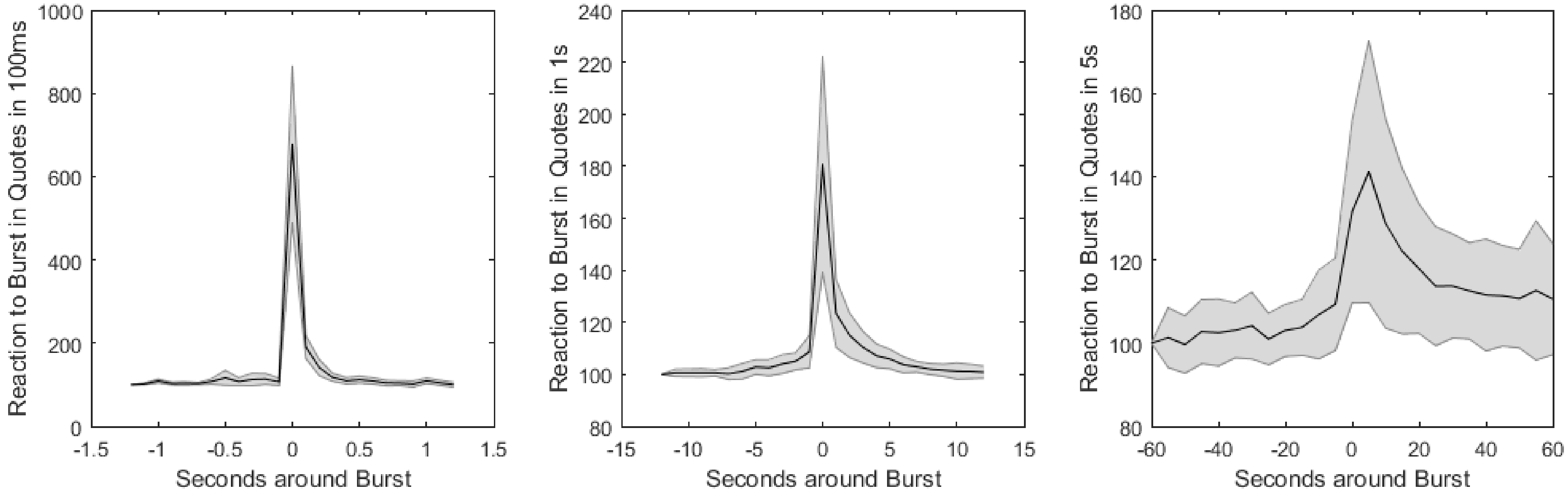

Figure 4.5: Return volatility around bursts in quotes.

This figure illustrates return volatility in a 2.5-second (25-second, 2-minute) window around bursts in quotes on the left (middle, right). The black line represents the average across all 100-millisecond (1-second, 5-second) intervals before, during, and after bursts, as well as across all stocks in our sample, where the initial value was normalized to 100. The shaded area represents one standard deviation from the mean.

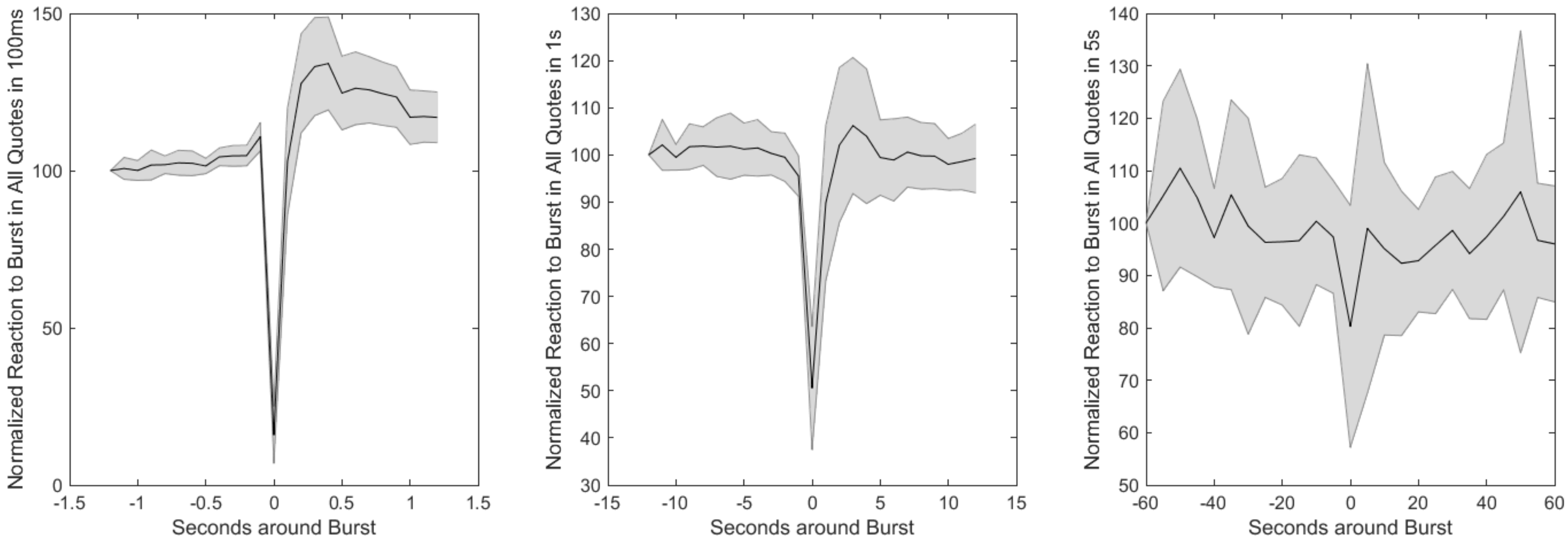

Figure 4.6: Average order size around bursts in quotes.

This figure studies average order size in a 2.5-second (25-second, 2-minute) window around bursts in quotes on the left (middle, right). The black line represents the average across all 100-millisecond (1-second, 5-second) intervals before, during, and after bursts, as well as across all stocks in our sample, where the initial value was normalized to 100. The shaded area represents one standard deviation from the mean.

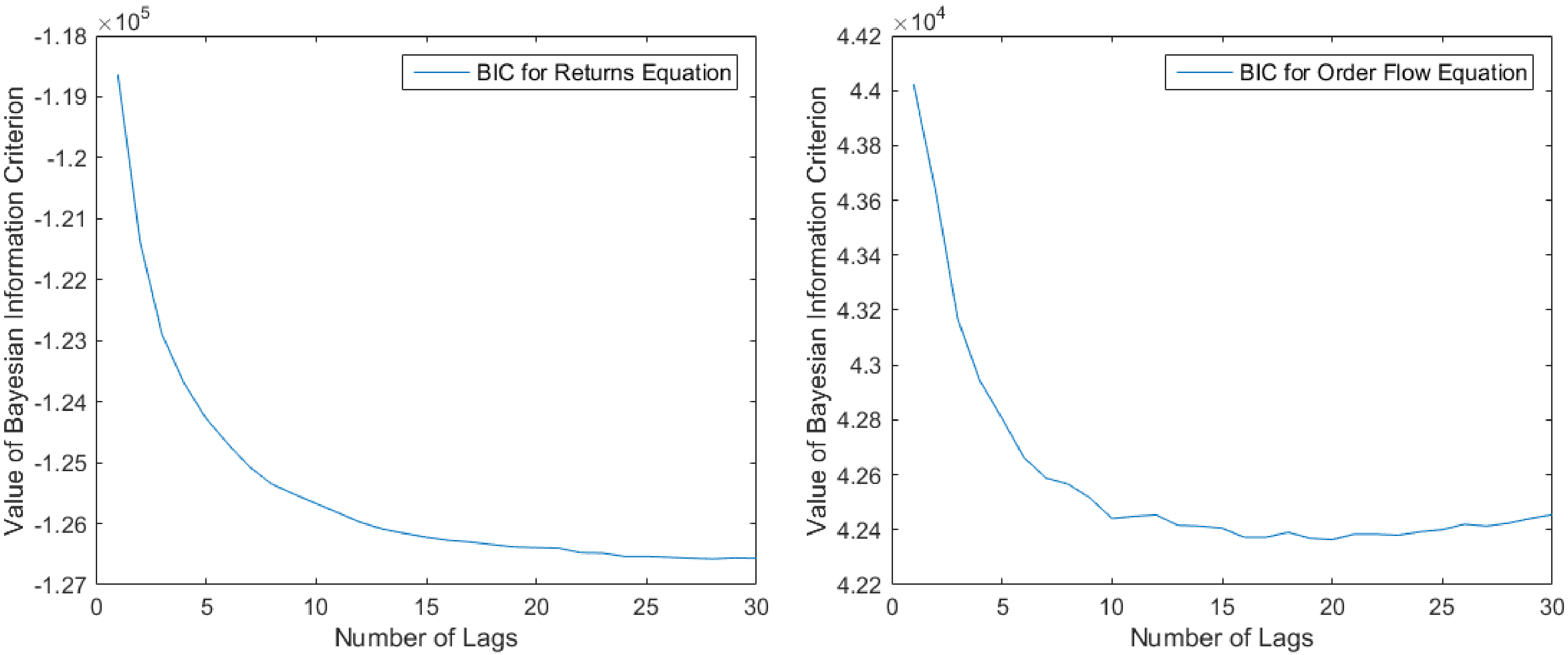

Figure 4.7: Lag Selection Using Bayesian Information Criterion.

This figure represents the values of the Bayesian information criterion for the returns and order flow equations for one trading day of AA stock. We obtained similar results for other stocks and days and when using the Akaike information criterion.

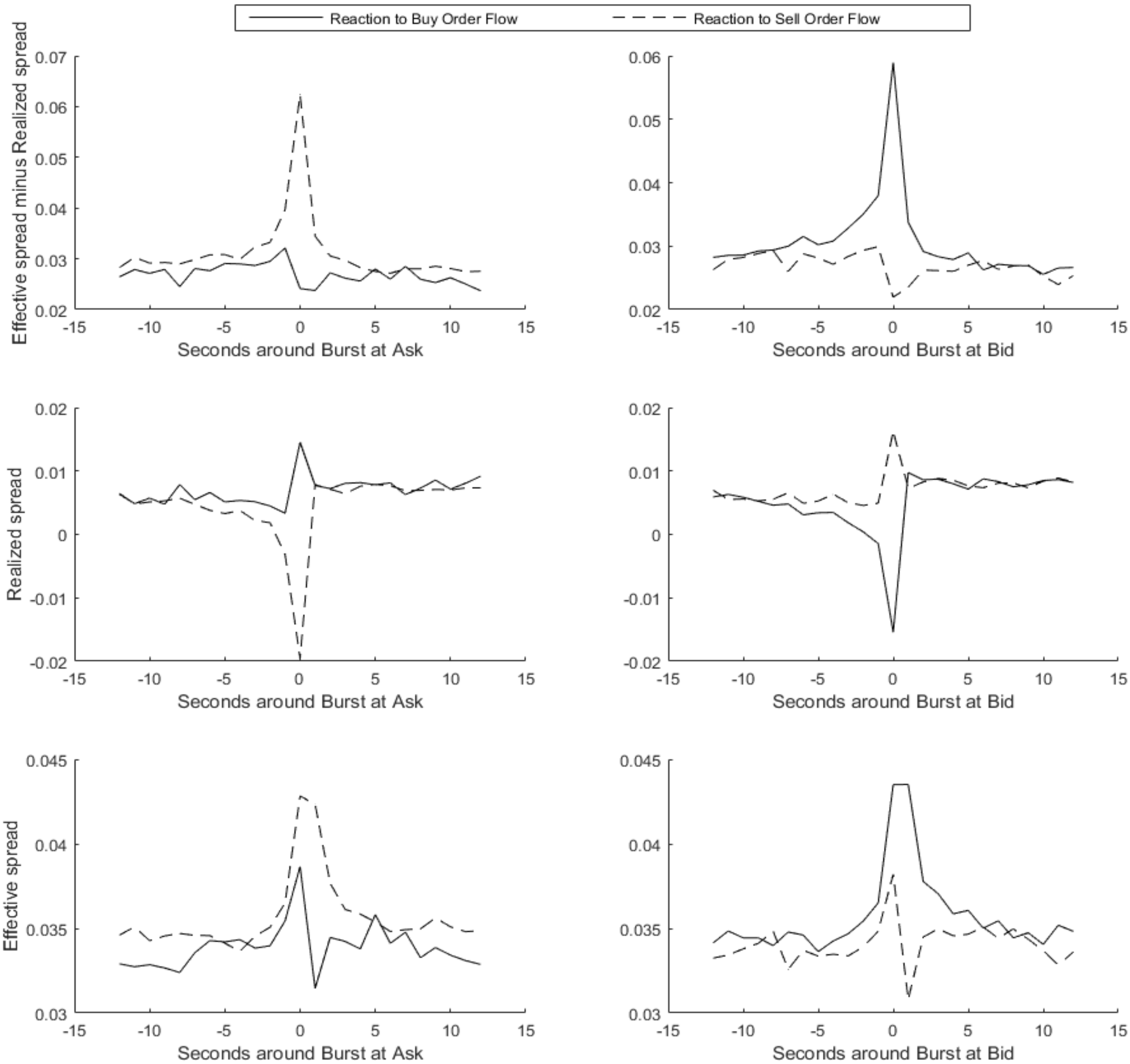

Figure 4.8: Adverse Selection, Realized and Effective Spread Around Bursts, One-Second Intervals.

This figure studies the adverse selection proxy in the top panel, the realized spread in the middle panel, and the effective spread in the bottom panel, before, during, and after bursts. The left panels show the reaction of the variables to a burst at ask with no concurrent burst at bid. The right panels show the same reaction to a burst at bid with no concurrent burst at ask. The reported values are percentage values of the mid-quote and averaged across 15 companies. They show the reaction of the variables to buy (sell) order flow in full (dashed) line. This figure is based on one-second intervals and a 99.9% threshold for bursts.

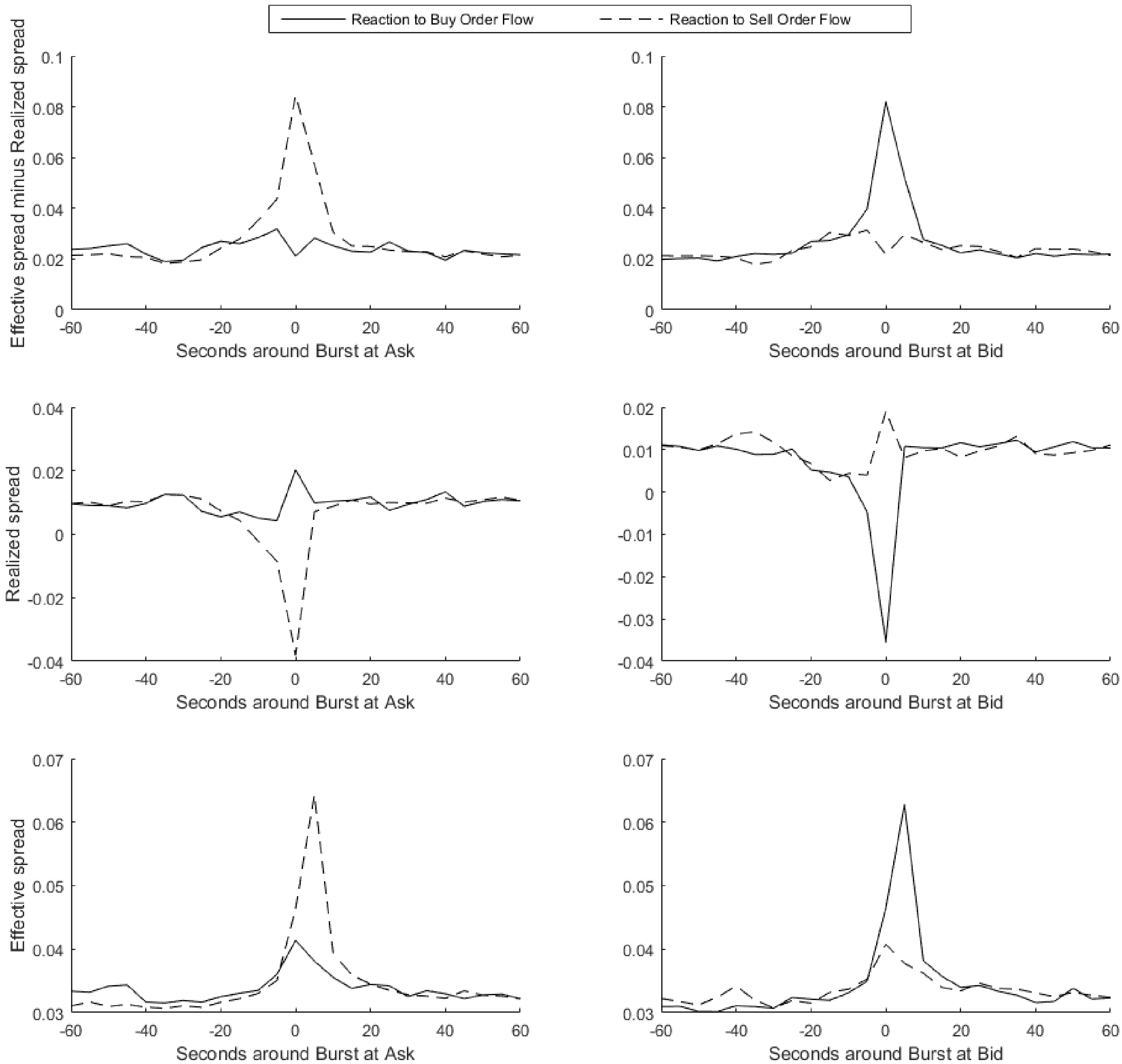

Figure 4.9: Adverse Selection, Realized and Effective Spread Around Bursts, Five-Second Intervals.

This figure studies the adverse selection proxy in the top panel, the realized spread in the middle panel, and the effective spread in the bottom panel, before, during, and after bursts. The left panels show the reaction of the variables to a burst at ask with no concurrent burst at bid. The right panels show the same reaction to a burst at bid with no concurrent burst at ask. The reported values are percentage values of the mid-quote and averaged across 15 companies. They show the reaction of the variables to buy (sell) order flow in full (dashed) line. This figure is based on five-second intervals and a 99.9% threshold for bursts.

Chapter 5

A Model of Price Impact and Market Maker Latency

Jakub Rojček[1]

Abstract

Price impact measures the difference between the best quoted price and the realized price as a function of order size. This paper analyzes how price impact depends on the latency that a market maker is subject to. I propose a tractable model which allows incorporating both order size and latency effects as determinants of price impact. The model is solved analytically and is novel in the theoretical microstructure literature. Larger latency increases adverse selection costs to the market maker and reduces his probability of trading with a slow investor. A larger order size decreases the slow trader's outside option, making him susceptible to accept a worse price for his trade. It is shown that the first-order effect of increased latency and increased order size is to increase price impact. Their joint impact is also positive. When the probability of trading is taken into consideration, the utility of the slow institutional investor decreases with latency.

Keywords: Price Impact, High-Frequency Trading, Trade Size, Latency, Market Quality, Welfare
JEL classification: D53 · G14 · C72.

5.1. Introduction

The joint influence of trading speed and order size has been largely neglected in recent high-frequency trading research. Most analyses have focused on understanding the impact of high-frequency trading on market quality and its influence on execution quality for traditional institutional investors in terms of the bid-ask spreads valid for unit size orders.[2] While theoretical models have shown in this setting

[1] I am grateful to Ramazan Gençay, Boyan Jovanovic, Albert Menkveld, and Alexandre Ziegler for providing me with comments, discussion and suggestions on this paper. All errors and omissions remain mine.

[2] For an overview of the literature on high-frequency trading, see the surveys by Jones (2013) and O'Hara (2015).

that algorithmic traders in continuous limit order markets derive a competitive advantage from faster analysis of order book evolution and other information,[3] the empirical literature has focused on the impact of algorithmic trading on traditional measures of market quality.[4] The majority of the empirical literature confirms the positive first-order effects of HFT, especially lower bid-ask spreads and faster price discovery. However, Hendershott et al. (2011) report that the presence of HFTs also decreases the depth of the order book and increases the costs of executing large orders. Taking execution costs into consideration, Tong (2013) measures the execution shortfall of institutional investors and finds that HFTs increase transaction costs for them. It is this overall effect of increased price impact which large institutional traders suffer that this paper focuses on. I present a model which incorporates the joint effect of latency and order size in a setting with a high-frequency market maker, high-frequency snipers, and slow investors. The model predicts that higher market maker latency and larger order size lead to higher price impact. Taking the probability of trading into consideration, slow institutional investors' utility is strictly deteriorating in larger order size and latency. The following sections present a review of the related literature, the model, comparative statics of the quoted price, and a welfare analysis.

5.2. Related Literature

The existing theoretical literature on HFT does not provide a model of price impact as a function of latency. There are two related streams of theoretical models. Repeated double auction models, which represent synchronous trading; and asynchronous trading models, which represent trading in continuous time. This section reviews the theoretical predictions of models in these two categories.

The double auction models of Roşu (2014), Rostek and Weretka (2015), Du and Zhu (2015), and Foucault et al. (2015) build on the models of Kyle (1985), Vayanos (1999), and Vives (2011).

Roşu (2014) models fast traders as those receiving information about the fundamental value instantly and slow traders as those receiving the information with a lag. Both categories are speculators. The comparative statics of price impact are derived as a function of the number of fast and slow traders, but it is not obvious how they would generalize to changing trading frequency. Moreover, it is not clear

[3] Foucault et al. (2015), Aït-Sahalia and Saglam (2014), Biais et al. (2015) and Hoffmann (2014).

[4] See Jovanovic and Menkveld (2011), Brogaard (2010), Hasbrouck and Saar (2013), Riordan and Storkenmaier (2012), Tong (2013), and Hendershott et al. (2011).

how the market maker's latency would be modelled. Rostek and Weretka (2015) show that for traders, maximizing welfare and stabilizing liquidity through disclosure of information about fundamentals at the same time represents a trade-off. The traders in their model balance the present execution value against future price impact. In a rational expectations equilibrium, traders split their orders optimally. The point from which I try to depart in my model is the synchronicity of traders arrivals. Du and Zhu (2014) depart from the synchronicity of arrivals in the double auction framework. They introduce one fast trader who is in the market every time step, while the rest of the traders arrives only every certain number of time steps. This leads to interesting results where the fast trader prefers higher frequency of trading, whereas the slow traders prefer slower trading. However, because this model falls into the category of double auction markets, traders do not suffer the latency effect per se. Their supplies change based on their frequency of trading, but not based on the risk they bear due to latency, during which they cannot change their orders. The present paper models this latency effect directly as a parameter of the price impact function. The closest to the modelling goal of the current paper is the paper by Foucault et al. (2015), where the speculator can receive the news about the fundamental value with a time advantage. This can be considered latency. In the repeated double auction setting they consider, the authors cannot solve for the equilibrium analytically and resolve to numerical solutions. This might lead one to conclude that double auction models with latency might be difficult to solve analytically. The current paper solves for price impact as a function of latency and size analytically. In order to achieve this, it uses the asynchronous arrivals modelling framework as a starting point.

Departing from the double auctions modelling framework are the asynchronous trading models of Menkveld and Zoican (2016), Budish et al. (2015), and Chacko et al. (2008).

Using Poisson arrivals, Menkveld and Zoican (2016) model liquidity traders (submitting market orders), HF Bandits (also submitting market orders), HF market makers (submitting limit orders), and good or bad news about the fundamental value. The net effect of latency on the spread depends on the news to liquidity traders ratio. What this model is lacking is the effect of trade size on prices. Budish et al. (2015) predict that in a limit order market with competing fast traders, the quoted size will always be one unit. An increasing presence of these fast snipers means that the liquidity provider can update his order only with diminishing probability in case it becomes mispriced. This results in a partial equilibrium, where the book contains a single unit limit order on each side of the market.

However, apart from this prediction, their model does not provide the price impact function as a function of size, because their equlibrium size is always one, nor as a function of latency. Chacko et al. (2008) model sell limit orders as writing a perpetual American call option, requiring delivery of the underlying block of shares upon execution. Similarly, a limit order to buy is like a short position in an American put option. What is specific about this model is that the limit orders have to be executed immediately in order to be able to use option pricing techniques. To ensure immediate execution, the initiator of a transaction offer (the option writer) must offer a price at which it is currently optimal for the receiver of the transaction offer (the option owner) to exercise the option early. In effect the market order is modelled as a limit order, which is submitted at the best opposite quote in order to ensure immediate execution. This structure does not permit an analysis of the effect of latency on price impact. What it allows, however, is to consider the impact of size, which is derived from the market maker having to resell the inventory back to the market. As a result, because the traders arrive at a frequency which decreases with order size, this directly translates into positive price impact, with its shape coming from the optimal execution rules for perpetual American options.

The goal of this paper is to obtain a tractable model of price impact as a function of order size and the market maker's latency. To achieve this goal, I propose a continuous time, asynchronous arrival model that consists of a risk-neutral monopolist market maker, a risk-averse buyer and seller, and high-frequency bandits. The buyer and the seller arrive in a stylized fashion according to a Poisson process and look for immediate trading opportunities. They submit a market order if the utility of doing so is at least their reservation utility. The reservation utility is the utility of submitting a limit order and waiting for a counterparty with opposite trading needs. In the general case, the fundamental value is modelled as a Brownian motion and it is possible to solve for the ask price numerically. I also provide a closed-form solution for price impact as a function of latency for the special case in which changes in the fundamental value over very short discrete intervals follow a uniform distribution. The paper provides novel insights on the role of high-frequency market makers and their contribution to liquidity and welfare in modern financial markets. It introduces a novel modelling framework in the theoretical microstructure literature providing scope for many surmisable extensions.

5.3. General Modelling Framework of Price Impact with Latency

This section presents a general framework that allows solving for the equilibrium ask price numerically. Further assumptions are used in the next section to derive an analytical expression for the ask price in order to facilitate the comparative statics and welfare analysis.

There is one financial asset, whose fundamental value at time t is denoted v_t. The fundamental value's dynamics is represented by a continuous-time stochastic process with zero drift and variance proportional to the time passed, $\sigma^2 t$, where the parameter σ is the volatility of the fundamental value. Let's denote the distribution of the increment $x := v_T - v_t$ as $F(x; \mu = 0, \sigma^2(T-t))$.

There is one buyer, one seller, and N risk-neutral high-frequency traders (HFT) in the market. One of the HFTs is currently the market maker (HFM), while the remaining $N-1$ HFTs are high-frequency bandits (HFB), who wait for mispriced limit orders and in that case, they snipe them and realize a profit.[5] The buyer and the seller are risk-averse.

The flow of events in the model is the following.[6] At time t the market maker submits a quote to sell, a, which is called the *ask price*. The ask price is expressed as a deviation from the current fundamental value v_t, so that the overall price the market maker sets for the asset is $v_t + a$. The market maker suffers a *latency* Δ and can return and update his ask price only after time Δ passes.[7] Two events can happen. Either the buyer enters according to a point process with intensity λ_B at a random time T_B before time Δ passes, evaluates whether he will demand immediacy and trade on the current ask, or one of the HFTs arrives after time Δ passes, in which case the market maker adjusts the ask if he arrives before one of the HFBs. If an HFB arrives before the market maker, the HFB snipes the ask if it is mispriced.[8]

The buyer has a private valuation of π_B in addition to the asset's current fundamental value. He will demand immediacy if his utility of holding the asset minus the ask price he pays is at least his reservation utility from submitting a limit order and waiting for the seller. If the buyer submits a

[5] This setting generalizes to a dynamic arrivals setting where buyers and sellers arrive according to the stylized fashion described in detail below.

[6] The corresponding flow of events for the closed-form example is summarized in Figure 5.1.

[7] Latency in a strict sense would mean that an ask set at time $t\Delta$ would be valid in the time interval $[(t+1)\Delta, (t+2)\Delta)$ based on information at time $(t-1)\Delta$. This is because processing information and quote submission are also subject to latency. For this model, I assume that the impact of latency can be simplified to the starting specification with the market maker's ask set at time $t\Delta$ being valid in interval $[(t)\Delta, (t+1)\Delta)$ based on the information at time $t\Delta$.

[8] The slow traders arrive in continuous time and the HFTs arrive at discrete points in time, Δ time units apart from each other.

market order, he will disappear from the market and the market maker will post a bid price at which he is willing to buy back the asset from the arriving seller and eliminate his inventory.

Then the seller arrives[9] according to a point process with intensity λ_S. The same reasoning as for the buyer applies. The seller decides whether to directly trade with the market maker or to submit a limit order and wait for the arrival of a new buyer, who will arrive at a random time T_{B2}. The seller submitting a limit order faces the risk that the fundamental value moves in an adverse direction. In that case, he might suffer a loss. On the other hand, if the fundamental value moves too much in a favourable direction, the buyer will not trade with the seller's limit order. The seller takes these possibilities into consideration when setting his ask price.

I assume that the traders trade Q shares at once and there is no uncertainty about this quantity.[10] As in Chacko et al. (2008), the arrival rate of traders on the opposite side of the market, in this case the seller, is assumed to be a decreasing function of the trade's size.

By submitting a limit order to sell at price a, the market maker issues an option to the fundamental buyer. The fundamental buyer executes on this limit order if $v_{T_B} - v_t - a + a' \geq 0$, where a' is the reservation ask. The reservation ask is the maximum price the buyer is willing to pay to trade the asset immediately and depends on the reservation value derived from submitting a buy limit order and waiting for a seller. The reservation ask price depends on future trading opportunities; it is computed in closed-form for the case of a uniform distribution in the next section. If the slow buyer does not arrive before time Δ passes and the market maker's limit order becomes mispriced, meaning that $v_{t+\Delta} > v_t + a$, the remaining $N-1$ HFBs will try to snipe it. Each one will be successful with equal probability $\frac{1}{N}$. With probability $\frac{1}{N}$, the market maker will be successful at cancelling this mispriced order.[11]

Were the market maker only facing the slow buyer and no HFBs, his payoff could be decomposed into a short position in an American call option plus owning a cash-or-nothing digital call option, both with strike price of $a - a'$ and the payoff of the cash-or-nothing call option equal to a'. The buyer on

[9] I assume that buyer and seller arrivals alternate in a deterministic fashion.

[10] This is a classical assumption in the microstructure literature (Ho and Stoll (1981)), which is relaxed in Budish et al. (2015) and particularly in the optimal execution literature.

[11] The corresponding payoff of the market maker in the closed-form example is depicted in Figure 5.2.

the other hand owns an American call option with the same strike price, $a - a'$. Formally, the market maker's payoff is

$$\mathbb{E}[LP(a)] = \mathbb{E}[(a - (v_{T_B} - v_t))\, \mathbb{1}_{v_{T_B} - v_t \geq a - a'} \mathbb{1}_{T_B \leq \Delta}]. \tag{5.1}$$

Assuming that the buyer arrives according to a Poisson process with intensity λ_B, this expectation can be written as

$$\mathbb{E}[LP(a)] = \int_0^\Delta \lambda_B e^{-\lambda_B y} \int_{a-a'}^\infty (a - x) F(\mathrm{d}x; 0, \sigma^2 y)\mathrm{d}y. \tag{5.2}$$

In the case with HFBs present in the market, the payoff of the market maker has three parts, one arising from trading with the slow buyer, the other two from trading with the HFBs. The market maker trades with the HFBs if the ask price at time $t + \Delta$ is mispriced ($v_{t+\Delta} > v_t + a$), provided that the slow trader does not arrive before $t+\Delta$ or if he does, he decides not to trade with the market maker. The market maker's payoff from trading with HFBs is thus equivalent to selling $\frac{N-1}{N}$ call options with strike price a. Equation (5.1) generalizes to

$$\begin{aligned} \mathbb{E}[LP(a)] &= \mathbb{E}[(a - (v_{T_B} - v_t))\, \mathbb{1}_{v_{T_B} - v_t \geq a - a'} \mathbb{1}_{T_B \leq \Delta}] \\ &- \frac{N-1}{N} \mathbb{E}[(v_{t+\Delta} - v_t - a)\, \mathbb{1}_{v_{t+\Delta} - v_t \geq a} \mathbb{1}_{T_B > \Delta}] \\ &- \frac{N-1}{N} \mathbb{E}[(v_{t+\Delta} - v_t - a)\, \mathbb{1}_{v_{t+\Delta} - v_t \geq a} \mathbb{1}_{T_B \leq \Delta} \mathbb{1}_{v_{T_B} - v_t < a - a'}], \end{aligned} \tag{5.3}$$

and the expectations can be computed as

$$\begin{aligned} \mathbb{E}[LP(a)] = &\int_0^\Delta \lambda_B e^{-\lambda_B y} \int_{a-a'}^\infty (a - x) F(\mathrm{d}x; 0, \sigma^2 y)\mathrm{d}y \\ &- \frac{N-1}{N} \left(\int_\Delta^\infty \lambda_B e^{-\lambda_B y} \mathrm{d}y \right) \int_a^\infty (x - a) F(\mathrm{d}x; 0, \sigma^2 \Delta) \\ &- \frac{N-1}{N} \int_0^\Delta \lambda_B e^{-\lambda_B y} \int_{-\infty}^{a-a'} \int_{a-x}^\infty (z + x - a) F(\mathrm{d}z; 0, \sigma^2(\Delta - y)) F(\mathrm{d}x; 0, \sigma^2 y)\mathrm{d}y, \end{aligned} \tag{5.4}$$

where the first line represents the possible profit by trading with the slow trader, the second line the loss due to the mispriced limit order being sniped by one of the HFBs at the end of the latency period if the slow buyer only arrives after Δ time passes, and the third line the case where the slow trader arrives before Δ time passes, decides not to trade because $v_{T_B} - v_t < a - a'$ and subsequently the fundamental value rises to $v_{t+\Delta} - v_t > a$.[12]

[12] The last term might be neglected if the probability $\int_0^\Delta \lambda_B e^{-\lambda_B y} \int_{-\infty}^{a-a'} \int_{a-x}^\infty F(\mathrm{d}z; 0, \sigma^2(\Delta - y)) F(\mathrm{d}x; 0, \sigma^2 y)\mathrm{d}y$ is small enough.

The reservation ask a' is the highest price the buyer is willing to pay to obtain the asset from the market maker. It is the price that equates the immediately available utility with the expected utility the buyer could obtain by submitting a buy limit order at price b and waiting for a seller.[13] The seller's private valuation for the asset is $-\pi_S$, where $\pi_S > 0$. The payoff structure is depicted in Figure 5.3, where the buyer's payoff is increasing in the fundamental value increment, but shrinks to zero once the fundamental value increment exceeds the seller's reservation value.[14] The expected payoff from submitting the limit order is the following

$$V_{B,LO}(b) = \mathbb{E}[u(\pi_B - v_{T_B} - b + v_{T_S})|v_{T_B} + b \geq v_{T_S} - \pi_S]. \tag{5.5}$$

The reservation ask price is then the price which solves the following equation at time T_B

$$u(\pi_B + v_{T_B} - a') = V_{B,LO}(b). \tag{5.6}$$

Given the reservation ask price, the market maker then uses Equation (5.1) to set a such that he fulfills an equilibrium condition. He might either set a in order to maximize his payoff or such that his expected payoff equals the expected HFB's payoff in case the competition from HFBs prevents profit maximizing behavior. One can solve for the equilibrium ask price a numerically in the general case. In the next section, I will solve an example which allows for a closed-form solution.

5.4. Closed-form Example

Generally, it is not possible to solve for the ask price in the above problem in closed-form. This section provides an example of a closed-form solution that allows investigating the role of latency in the price impact function.

The fundamental value's dynamics is now represented by a discrete-time stochastic process. As before, we let Δ denote the market maker's latency. We assume that the change in the fundamental value over the short interval Δ is distributed according to a uniform distribution $\mathcal{U}(-\sqrt{3}\sigma\sqrt{\Delta}, \sqrt{3}\sigma\sqrt{\Delta})$.[15]

[13] The bid price, b, is again expressed as a deviation from the fundamental value at time T_B and the overall bid price is $v_{T_B} + b$. The buyer sets b in order to maximize $V_{B,LO}(b)$.

[14] We assume that the seller accepts the limit order if his payoff is non-negative. However, he could also optimize and submit a sell limit order, in which case we assume that the previous buyer exits the game. The number of optimizing agents is driven by computational considerations and trades off the precision with computation time.

[15] Empirically, high-frequency returns are not normally distributed and have heavy tails with large spikes around zero.

The expected change in the fundamental value is thus zero and the variance of the fundamental value change is proportional to the latency, $\sigma^2\Delta$.

The flow of events is summarized in Figure 5.1. A fundamental buyer arrives according to a Poisson point process with intensity λ. The probability of n buyers arriving by time Δ is $\mathbb{P}[B(0,\Delta] = n] = \frac{(\lambda\Delta)^n}{n!}e^{-\lambda\Delta}$. We assume that the latency is sufficiently small that two and more arrivals of fundamental traders are very unlikely during the Δ interval. The probability that exactly one fundamental buyer arrives by time Δ is $\lambda\Delta + \mathcal{O}(\Delta^2)$, which comes from applying a Taylor approximation to the probability of Poisson arrivals. The probability of no buyer arriving by time Δ is then $1 - \lambda\Delta + \mathcal{O}(\Delta^2)$.

In order to be able to obtain a closed-form solution to the market maker's problem, we also suppose that the fundamental buyer faces the fundamental value at the end of the interval Δ. This means that if the buyer arrives during $(t, t+\Delta]$, the fundamental value he takes into account is $v_{t+\Delta}$.[16] When the buyer arrives, he can execute on the current ask price set by the market maker, a, or submit his own limit order and wait for the potential fundamental seller. In addition to the fundamental value $v_{t+\Delta}$, the buyer derives a private value π_B from holding the asset. Let a' denote the buyer's *reservation ask*. The reservation ask is the maximum price the buyer is willing to pay to trade the asset immediately and depends on the reservation value derived from submitting a buy limit order and waiting for a seller. The buyer's decision whether to execute or not based on the current ask price, a, is depicted in Figure 5.2 and can be summarized as follows

$$v_{t+\Delta} - v_t \begin{cases} \geq a - a' & \text{buyer executes,} \\ < a - a' & \text{buyer does not execute.} \end{cases} \tag{5.7}$$

Market Maker's Problem

Because the stylized limit order book can only hold one ask, only one HFT can become a market maker (HFM). By contrast with Section 5.3, I assume that the slow buyer is never successful at picking off a mispriced ask, so his payoff lies between 0 and a'. If the ask becomes mispriced, it is cancelled by

[16] Weighting the market maker's payoff by the arrival time leads to fixed point problems from which it is not possible to back out the ask price as a function of parameters and basic functions in closed form.

the HFM with probability $\frac{1}{N}$ and picked off by one of the HFBs with probability $\frac{N-1}{N}$. The expected payoff to the market maker is given by

$$\mathbb{E}[LP(a)] = \lambda\Delta \int_{a-a'}^{a} (a-x)\frac{1}{2\sigma\sqrt{\Delta}\sqrt{3}}\mathrm{d}x - \frac{N-1}{N}\int_{a}^{\sigma\sqrt{\Delta}\sqrt{3}} (x-a)\frac{1}{2\sigma\sqrt{\Delta}\sqrt{3}}\mathrm{d}x \tag{5.8}$$

$$= \lambda\Delta\frac{a'^2}{4\sqrt{3}\sqrt{\Delta}\sigma} - \frac{N-1}{N}\left(\frac{a^2}{4\sqrt{3}\sqrt{\Delta}\sigma} - \frac{a}{2} + \frac{1}{4}\sqrt{3}\sqrt{\Delta}\sigma\right). \tag{5.9}$$

A HFB has a chance of $\frac{1}{N}$ that he would successfully snipe a mispriced ask. His expected payoff is given by

$$\mathbb{E}[SP(a)] = \frac{1}{N}\int_{a}^{\sigma\sqrt{\Delta}\sqrt{3}} (x-a)\frac{1}{2\sigma\sqrt{\Delta}\sqrt{3}}\mathrm{d}x \tag{5.10}$$

$$= \frac{1}{N}\left(\frac{a^2}{4\sqrt{3}\sqrt{\Delta}\sigma} - \frac{a}{2} + \frac{1}{4}\sqrt{3}\sqrt{\Delta}\sigma\right). \tag{5.11}$$

As in Menkveld and Zoican (2016), our equilibrium condition states that the expected payoff of the HFM and HFBs must be equal

$$\mathbb{E}[LP(a)] = \mathbb{E}[SP(a)]. \tag{5.12}$$

By applying this condition and rearranging terms, we obtain the following quadratic equation in a, which must hold

$$-\frac{a^2}{4\sqrt{3}\sqrt{\Delta}\sigma} + \frac{a}{2} - \frac{1}{4}\sqrt{3}\sqrt{\Delta}\sigma + \lambda\Delta\frac{a'^2}{4\sqrt{3}\sqrt{\Delta}\sigma} = 0. \tag{5.13}$$

This equation has two solutions

$$a^*_{1,2} = \sqrt{3}\sqrt{\Delta}\sigma \pm \sqrt{\lambda\Delta}a'. \tag{5.14}$$

Economically meaningful is the solution which increases in the reservation ask price, $a^* = \sqrt{3}\sqrt{\Delta}\sigma + a'\sqrt{\lambda\Delta}$. The first term in the equilibrium ask price comes from the snipe off part. It represents the ask price which equates the expected loss of the market maker with the HFB's expected profit in the case that the probability of the slow buyer's arrival is zero. The second term represents an adjustment for the expected profit from trading with the slow buyer.

Proposition 1. Market maker's ask price. *Let there be N high-frequency traders in the market. Given the latency Δ, the buyer's arrival rate λ, and assuming $v_\Delta - v_0 \sim \mathcal{U}(-\sqrt{3}\sqrt{\Delta}\sigma, \sqrt{3}\sqrt{\Delta}\sigma)$, the equilibrium ask price is*

$$a^* = \sqrt{3}\sqrt{\Delta}\sigma + \sqrt{\lambda\Delta}a'. \tag{5.15}$$

Proof. Follows from the steps above. $\square$

We next solve for the highest reservation price a' that a buyer is willing to pay.

Buyer's Reservation Value

I assume that both the buyer and the seller are risk averse and have exponential constant absolute risk-aversion utility functions $u(x) = 1 - e^{-\alpha x}$, where α is the risk-aversion coefficient.

The buyer arrives at time $t + \Delta$ and observes the fundamental value at that time, $v_{t+\Delta}$. He executes if the value of submitting a market order $1 - e^{-\alpha(\pi_B + v_{t+\Delta} - v_t - a)}$ exceeds than the value of submitting a limit order, which we compute below. Otherwise, the buyer submits a limit order and waits for a seller, who arrives according to a Poisson process with intensity $\lambda_S(Q)$. The arrival intensity is a decreasing function of the order size Q, meaning that the buyer would in expectation have to wait longer for an opposite side trader if his order is larger. Although this paper's findings do not depend on the precise functional form, for illustration purposes we use the functional form proposed by Chacko et al. (2008), where the intensity is inversely related to the quantity traded, $\lambda_S(Q) = \frac{\Lambda_S}{Q}$. The parameter Λ_S represents the arrival intensity of unit size order seller. This is equivalent to assuming that the demand for trading is stationary per unit of time as in Garman (1976). I will use λ_S and $\lambda_S(Q)$ interchangeably.

It is assumed that the mean arrival time of the seller, $\frac{1}{\lambda_S}$, is much larger than the market maker's latency, Δ. Because the innovations to the fundamental value are uniformly distributed with mean zero and variance $\sigma^2\Delta$, it follows from the central limit theorem for sums that the sum of such innovations, $\sum_{j=1}^{J}(v_{t+j\Delta} - v_{t+(j-1)\Delta})$ is normally distributed with mean zero and variance $\sigma^2 J\Delta$, for large values of J. The change in the fundamental value between the buyer's arrival (reset to 0 for convenience) and the seller's arrival time T_S is approximately normally distributed with mean 0 and variance $\sigma^2 T_S$, $v_{T_S} \sim N(0, \sigma^2 T_S)$. This simplifies the calculations for the buyer's reservation value and enables closed-form solution for the buyer's reservation ask price.

Suppose that the buyer submits a limit order priced at the fundamental value v_t.[17] It is assumed that the seller will execute on the buyer's order in case his payoff is not negative. His utility from the trade is $1 - e^{-\alpha(\pi_S - v_{T_s})}$, where $-\pi_S$ is his private valuation for the asset in addition to its current fundamental value. It thus follows that the seller executes on the buyer's order in case the fundamental value is below π_S, as this still leaves the seller with a positive trading surplus. The buyer's payoff increases up until this point and is zero once the fundamental value is larger than π_S. The seller's and buyer's payoffs from trading are depicted in Figure 5.3. The seller's decision can be summarized as follows

$$v_{T_S} \begin{cases} \leq \pi_S & \text{seller executes,} \\ > \pi_S & \text{seller does not execute.} \end{cases} \tag{5.16}$$

Taking the seller's decision into consideration, the following lemma states the reservation value of the buyer.

Lemma 5.4.1. Buyer's reservation value. *Given the seller's arrival rate λ_S and private valuation π_S, the buyer's risk-aversion coefficient α and private valuation π_B, the fundamental value's volatility σ and $\lambda_S > \frac{\alpha^2\sigma^2}{2}$, the buyer's reservation value, $V_{B,LO}$, is the following:*

$$V_{B,LO} = 1 - \frac{\lambda_S e^{-\alpha\pi_B}}{\lambda_S - \frac{\alpha^2\sigma^2}{2}} \left(1 + e^{-\alpha\pi_S} e^{-\sqrt{\frac{2\lambda_S\pi_S}{\sigma^2}}} \left[\frac{\alpha\sigma}{2}\sqrt{\frac{\pi}{\lambda_S}} - \sqrt{\frac{\pi}{2}}\sqrt{\pi_S}\right]\right). \tag{5.17}$$

Proof. The proof is given in appendix A. □

In case the seller's private valuation π_S is much larger than the variance $\sigma^2 T_S$, the above expression is approximately equal to

$$V_{B,LO} \approx 1 - \frac{\lambda_S e^{-\alpha\pi_B}}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}. \tag{5.18}$$

In the following, the buyer's reservation utility is transformed to the maximum ask price a' at which he is willing to buy the asset from the market maker.

The Buyer's Reservation Ask Price

[17] v_t serves here as a reference point, the analysis and conclusions do not change due to this choice. It is equivalent to setting the bid, b, in Equation (5.5) to 0. This paper also does not aim at explicitly modelling the limit order submission decision of the buyer, who is being considered here as a liquidity trader with a binary choice.

The buyer will be indifferent between trading at the market maker's ask price and submitting a limit order if his utility of submitting the market order, $V_{B,MO}(a')$, equals his utility from submitting a limit order, $V_{B,LO}$, which we derived above as the buyer's reservation utility. The *reservation ask price* a' is the ask price that equates these two utilities. Using Equation (5.18) and $V_{B,MO}(a') = 1 - e^{-\alpha(\pi_B - a')}$ yields

$$1 - \frac{\lambda_S e^{-\alpha\pi_B}}{\lambda_S - \frac{\alpha^2\sigma^2}{2}} = 1 - e^{-\alpha(\pi_B - a')}. \tag{5.19}$$

Simplifying yields

$$e^{-\alpha\pi_B}\left(e^{\alpha a'} - \frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right) = 0. \tag{5.20}$$

Solving this equation for a' results in the following lemma:

Lemma 5.4.2. Reservation ask price. *Given the seller's arrival rate λ_S, the buyer's risk-aversion coefficient α and the fundamental value's volatility σ, the highest price at which the buyer is willing to buy the asset, a', is:*

$$a' = \frac{1}{\alpha}\log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right) \tag{5.21}$$

Proof. The proof follows from the steps above.[18] □

The Market Maker's Ask Price

The last step to compute the market maker's ask price is to insert the expression for the reservation ask price from lemma 5.4.2 into the general solution for the ask price given in proposition 1. This yields

Proposition 2. Market maker's ask price with reservation ask. *Let there be N high-frequency traders in the market. Given the latency Δ, the buyer's arrival rate λ, the seller's arrival rate λ_S and assuming $v_\Delta - v_0 \sim \mathcal{U}(-\sqrt{3}\sqrt{\Delta}\sigma, \sqrt{3}\sqrt{\Delta}\sigma)$ and that $\frac{1}{\lambda_S} >> \Delta$, the equilibrium ask price is*

$$a^* = \sqrt{3}\sqrt{\Delta}\sigma + \sqrt{\lambda\Delta}\frac{1}{\alpha}\log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right). \tag{5.22}$$

Proof. Follows from the steps above. □

In the following section we use this result to investigate the relationship market maker latency and price impact.

[18] The $\log(\cdot)$ represents the natural logarithm function.

5.5. Comparative Statics

This paper's main result is stated in proposition 3. It provides the comparative statics analysis of the ask price with respect to the model primitives.

Proposition 3. Price impact and latency. *Given the market maker's ask price a^* as stated in Equation (5.22), the ask price*

1. *increases in the market maker's latency Δ,*
2. *decreases in the seller's arrival rate λ_S,*
3. *increases in the trade size Q,*
4. *increases in the asset's volatility σ.*

Proof. The proof is outlined below. □

Let us first analyze the impact of latency Δ, starting from Equation (5.22) by taking derivatives with respect to Δ:

$$\frac{\partial a^*}{\partial \Delta} = \frac{\partial}{\partial \Delta}\left(\sqrt{3}\sqrt{\Delta}\sigma + \sqrt{\lambda\Delta}\frac{1}{\alpha}\log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)\right) \tag{5.23}$$

$$= \frac{\sigma}{2}\sqrt{\frac{3}{\Delta}} + \frac{1}{2}\sqrt{\frac{\lambda}{\Delta}}\frac{1}{\alpha}\log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right) > 0. \tag{5.24}$$

The first term comes from the increased adverse selection the market maker is facing and is positive. Thus, larger latency leads to higher adverse selection costs, which are compensated by a higher spread. The second term represents the mark up the market maker is able to charge the buyer on average due to the buyer's preference for immediacy. That term is also positive as the reservation ask price is positive.

The impact of the intensity of seller arrivals is determined in a similar fashion:

$$\frac{\partial a^*}{\partial \lambda_S} = \frac{\partial}{\partial \lambda_S}\left(\sqrt{3}\sqrt{\Delta}\sigma + \sqrt{\lambda\Delta}\frac{1}{\alpha}\log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)\right) \tag{5.25}$$

$$= \sqrt{\lambda\Delta}\frac{1}{\alpha}\frac{\lambda_S - \frac{\alpha^2\sigma^2}{2}}{\lambda_S}\frac{\lambda_S - \frac{\alpha^2\sigma^2}{2} - \lambda_S}{(\lambda_S - \frac{\alpha^2\sigma^2}{2})^2} \tag{5.26}$$

$$= -\sqrt{\lambda\Delta}\frac{1}{\alpha}\frac{\frac{\alpha^2\sigma^2}{2}}{\lambda_S(\lambda_S - \frac{\alpha^2\sigma^2}{2})} < 0. \tag{5.27}$$

The negative sign confirms the intuition that if the buyer's outside option is more valuable, because his chance of meeting a seller sooner is higher, he will be less willing to pay a high ask price.

The impact of trade size on the ask price can be obtained by recalling that $\lambda_S(Q) = \frac{\Lambda_S}{Q}$ and applying the chain rule,

$$\frac{\partial a^*(\lambda_S(Q))}{\partial Q} = \frac{\partial a^*}{\partial \lambda_S}\frac{\partial \lambda_S}{\partial Q} \tag{5.28}$$

$$= -\sqrt{\lambda\Delta}\frac{1}{\alpha}\frac{\frac{\alpha^2\sigma^2}{2}}{\lambda_S(\lambda_S - \frac{\alpha^2\sigma^2}{2})}\left(-\frac{\Lambda_S}{Q^2}\right) \tag{5.29}$$

$$= \sqrt{\lambda\Delta}\frac{1}{\alpha}\frac{\frac{\alpha^2\sigma^2}{2}}{\lambda_S(\lambda_S - \frac{\alpha^2\sigma^2}{2})}\left(\frac{\Lambda_S}{Q^2}\right) > 0. \tag{5.30}$$

The larger the quantity the buyer wants to trade, the longer he would need to wait for a potential seller, the lower is the value of his outside option and thus the higher is the price the market maker is able to charge.

The cross-derivative of the ask price with respect to latency Δ and the order size Q is given by

$$\frac{\partial a^*}{\partial Q \partial \Delta} = \frac{1}{2}\sqrt{\frac{\lambda}{\Delta}}\frac{1}{\alpha}\frac{\frac{\alpha^2\sigma^2}{2}}{\lambda_S(\lambda_S - \frac{\alpha^2\sigma^2}{2})}\left(\frac{\Lambda_S}{Q^2}\right) > 0. \tag{5.31}$$

The higher is the probability that a buyer with a lower valued outside option will come, the higher the ask price that the market maker can charge.

The effect of an increase in the volatility of the fundamental value on the ask price is:

$$\frac{\partial a^*}{\partial \sigma} = \frac{\partial}{\partial \sigma}\left(\sqrt{3}\sqrt{\Delta}\sigma + \sqrt{\lambda\Delta}\frac{1}{\alpha}\log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)\right) \tag{5.32}$$

$$= \sqrt{3}\sqrt{\Delta} + \sqrt{\lambda\Delta}\frac{1}{\alpha}\frac{\lambda_S - \frac{\alpha^2\sigma^2}{2}}{\lambda_S}\frac{\lambda_S\alpha^2\sigma}{(\lambda_S - \frac{\alpha^2\sigma^2}{2})^2} \tag{5.33}$$

$$= \sqrt{3}\sqrt{\Delta} + \sqrt{\lambda\Delta}\frac{\alpha\sigma}{\lambda_S - \frac{\alpha^2\sigma^2}{2}} > 0. \tag{5.34}$$

Volatility increases the ask price through two channels. First, a higher volatility leads to higher adverse selection costs for the market maker. Second, it reduces the value of the risk-averse buyer's outside option, allowing the market maker to charge a higher ask price.

Illustrative examples are provided in Figure 5.4. This figure represents the sensitivity of the ask price a^* to the latency Δ, order size Q, and volatility σ. The base parameters chosen are $\sigma = 0.2$, $\lambda = 1$,

$\Lambda_S = 10$, and $Q = 1$. The left panel shows the impact of latency and order size. The middle panel shows the impact of latency and volatility and the right panel the impact of order size and volatility. The ask price is increasing in latency and this increase is more prominent, the larger the order size. The ask price also increases in the volatility of the fundamental value.

5.6. Welfare Analysis

This section analyzes comparative statics of slow traders' and high-frequency traders' expected profits as well as the probability of trading.

High-frequency Traders' Profit

The profit of the HFM has to be equal in expectation to the profit of the HFBs. HFTs' profit is obtained by plugging the solution for the ask price from Equation (5.22) into the HFBs' profit Equation (5.10). The sensitivity of HFBs' profit to the model parameters is given in the following lemma.

Lemma 5.6.1. Equilibrium HFT profits. *Let there be N HFTs in the market. Given the seller's arrival rate λ_S, the buyer's risk-aversion coefficient α, the market maket's latency Δ and the fundamental value's volatility σ, the HFT's expected profit is*

$$\mathbb{E}[SP(a^*)] = \frac{1}{N}\left(\frac{\sqrt{\Delta}\lambda a'^2}{4\sqrt{3}\sigma}\right) \tag{5.35}$$

$$= \frac{1}{N}\left(\frac{\sqrt{\Delta}\lambda\left[\log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)\right]^2}{4\sqrt{3}\sigma\alpha^2}\right). \tag{5.36}$$

HFTs' profit

1. *decreases in the number of HFTs N,*
2. *increases in the buyer's arrival rate λ,*
3. *decreases in latency Δ,*
4. *decreases in the seller's arrival rate λ_S,*
5. *increases in the asset's volatility σ.*

Proof. The proof is outlined below. □

The effect of an increase in latency can be directly observed in Equation (5.35), where the increase in latency leads to an increase in HFT profits in proportion to $\sqrt{\Delta}$. This is due to the increased probability of the arrival of the slow buyer by time Δ; thus, it is more natural to normalize the profit by time. Doing so yields $\frac{1}{N}\left(\frac{\lambda a'^2}{4\sqrt{\Delta}\sqrt{3}\sigma}\right)$, which is decreasing in Δ. This is because as Δ rises, the dispersion of the fundamental value increases, lowering the chance that the fundamental value will lie in the execution interval $[a^* - a', a^*]$.

HFTs' profit increases in the arrival rate of the slow buyer by increasing the chance of a trade in the next Δ time interval. Profits decrease in the number of HFTs N, as they are divided by a larger number of possible liquidity providers. HFTs' profit also decreases in the seller's arrival rate. The reason is that a higher λ_S increases the buyer's reservation value:

$$\frac{\partial \mathbb{E}[SP(a^*)]}{\partial \lambda_S} = -\frac{1}{N}\left(\frac{\sqrt{\Delta}\lambda\sigma \log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)}{\sqrt{3}\lambda_S(\lambda_S - \frac{\alpha^2\sigma^2}{2})}\right) < 0. \tag{5.37}$$

The impact of volatility on HFTs' profit is positive. The profit increases if

$$\frac{\partial \mathbb{E}[SP(a^*)]}{\partial \sigma} = \frac{1}{N}\left(\frac{\sqrt{\Delta}\lambda \log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)}{2\sqrt{3}(\lambda_S - \frac{\alpha^2\sigma^2}{2})} - \frac{\sqrt{\Delta}\lambda\Big[\log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)\Big]^2}{4\sqrt{3}\alpha^2\sigma^2}\right), \tag{5.38}$$

which is positive provided that

$$\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}} - 1 > \frac{1}{4}\log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right). \tag{5.39}$$

Remembering that $\lambda_S > \frac{\alpha^2\sigma^2}{2}$ and setting $y = \frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}} > 1$, the last inequality is always true, because the function $y - 1 - \frac{1}{4}\log y$ is always positive for $y > 1$.

The effects of latency, order size, and volatility on HFTs' profit are illustrated in Figure 5.5. The left panel shows that HFTs' profit decreases in latency. The middle panel shows the positive effect of the volatility and the right panel the positive impact of the order size on profits.

Slow Buyer's Utility and the Probability of Trading

In the following, I compute the expected utility of the slow buyer, V_B. This is in general given by the buyer's utility from submitting a market order in case he executes on the market maker's ask quote and by the utility he derives from submitting a limit order in case (i) the ask price was higher than his reservation ask, or (ii) the high fundamental value created an arbitrage opportunity for the HFBs from which the slow buyer cannot profit:

$$V_B = \mathbb{E}\left[\left(1 - e^{-\alpha(\pi_B + v_\Delta - a^*)}\right) \mathbb{1}_{a^* - a' \leq v_\Delta \leq a^*}\right] + V_{B,LO}\mathbb{P}[a^* - a' > v_\Delta \vee v_\Delta > a^*]. \tag{5.40}$$

Before computing the buyer's expected utility, it is useful to investigate the drivers of the probability that the trade will happen, $\mathbb{P}[a^* - a' \leq v_\Delta \leq a^*]$.

Lemma 5.6.2. Equilibrium Probability of Trading. *Given the seller's arrival rate λ_S, the buyer's risk-aversion coefficient α and arrival rate λ, latency Δ and the fundamental value's volatility σ, the probability that a trade between the buyer and the market maker will happen is*

$$\begin{aligned}
\mathbb{P}[a^* - a' \leq v_\Delta \leq a^*] &= \lambda\Delta \int_{a^*-a'}^{a^*} \frac{1}{2\sqrt{3}\sqrt{\Delta}\sigma}\, dx &(5.41)\\
&= \lambda\sqrt{\Delta}\frac{a'}{2\sqrt{3}\sigma} &(5.42)\\
&= \frac{\lambda\sqrt{\Delta}\log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)}{2\sqrt{3}\sigma\alpha}. &(5.43)
\end{aligned}$$

The probability of trading

1. *decreases in the seller's arrival rate λ_S,*
2. *decreases in latency Δ,*
3. *increases in the buyer's arrival rate λ,*
4. *increases in the asset's volatility σ.*

Proof. The remainder of the proof is outlined below. □

From the calculation above it can be seen that the probability of trading between the market maker and the buyer does not directly depend on the market maker's ask price a^*. However, it depends on the reservation ask a'. The higher the seller's arrival rate, the lower the reservation ask price and the lower therefore the probability of trading with the market maker:

$$\frac{\partial\mathbb{P}[\cdot]}{\partial\lambda_S} = -\frac{\alpha\sigma\lambda\sqrt{\Delta}}{4\sqrt{3}\lambda_S(\lambda_S - \frac{\alpha^2\sigma^2}{2})} < 0. \tag{5.44}$$

Increasing the latency or the buyer's arrival rate increases the chance that the buyer will arrive by time Δ. On the other hand, increasing latency leads to a higher dispersion of the fundamental value, reducing the chance that it will fall in the acceptable trading range $[a^* - a', a^*]$, lowering the trading probability. We are interested in the effect per unit of time. Overall, this will be negative. Indeed,

$$\frac{\partial(\mathbb{P}[\cdot]/\Delta)}{\partial\Delta} = -\frac{a'}{4\sqrt{3}\Delta^{\frac{3}{2}}\sigma} < 0. \tag{5.45}$$

The impact of asset price volatility on the probability of trading is given by

$$\frac{\partial\mathbb{P}[\cdot]}{\partial\sigma} = \lambda\sqrt{\Delta}\frac{\frac{\frac{\alpha^2\sigma^2}{2}}{\lambda_S - \frac{\alpha^2\sigma^2}{2}} - \frac{1}{2}\log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)}{\sqrt{3}\alpha}. \tag{5.46}$$

This expression is positive provided that

$$\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}} - 1 > \frac{1}{2}\log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right) \tag{5.47}$$

Setting $y = \frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}} > 1$, this condition is always met, because the function $y - 1 - \frac{1}{2}\log y$ is always positive for $y > 1$.

The effects of latency, order size, and volatility on the probability of trading are illustrated in Figure 5.6. The left panel shows that higher latency reduces the probability of trading per unit of time. The middle panel shows the positive impact of the volatility of the fundamental value and the right the positive impact of the order size.

We are now in a position to compute the expected utility of the slow buyer. The expected utility can be divided into two components. The first is the payoff obtained from executing on the market maker's ask price. The second arises from submitting a limit order in case trading with the market maker is no longer the best option for the buyer.

Proposition 4. Equilibrium Slow Trader's Utility. *Given the seller's arrival rate λ_S, the buyer's arrival rate λ, his risk-aversion coefficient α and private valuation π_B, latency Δ and the fundamental value's volatility σ, the slow buyer's expected utility derived from a limit order is*

$$V_{B,LO}\mathbb{P}[a^* - a' > v_\Delta \vee v_\Delta > a^*] = \left(1 - \frac{\lambda_S e^{-\alpha\pi_B}}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)\lambda\Delta\left(1 - \frac{\log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)}{2\sqrt{3}\sqrt{\Delta}\sigma\alpha}\right). \tag{5.48}$$

The limit order payoff

1. *increases in the seller's arrival rate λ_S if $V_{B,LO}$ is sufficiently large,*

2. *increases in the latency Δ,*

3. *increases in the buyer's arrival rate λ.*

The slow buyer's expected utility derived from submitting a market order $V_{B,MO}$ is

$$\mathbb{E}\left[\left(1-e^{-\alpha(\pi_B+v_\Delta-a^*)}\right)\mathbb{1}_{a^*-a'\leq v_\Delta\leq a^*}\right]=\lambda\sqrt{\Delta}\frac{\log\left(\frac{\lambda_S}{\lambda_S-\frac{\alpha^2\sigma^2}{2}}\right)+e^{-\alpha\pi_B}\left(1-\frac{\lambda_S}{\lambda_S-\frac{\alpha^2\sigma^2}{2}}\right)}{2\sqrt{3}\alpha\sigma}. \tag{5.49}$$

The market order payoff

1. *increases in the buyer's arrival rate λ,*

2. *decreases in the seller's arrival rate λ_S if $V_{B,LO}>0$,*

3. *decreases in the latency Δ if $e^{\alpha\pi_B}>\frac{\frac{\partial a'}{\partial\sigma^2}}{\frac{a'}{\sigma^2}}$.*

The slow buyer's overall payoff $V_{B,LO}\mathbb{P}[a^*-a'>v_\Delta\vee v_\Delta>a^*]+V_{B,MO}$

1. *increases in the seller's arrival rate λ_S if $2\sqrt{3}\sigma\sqrt{\Delta}>a'$,*

2. *increases in the buyer's arrival rate λ,*

3. *decreases in latency Δ.*

Proof. The proof of Equation (5.48) and Equation (5.49) is given in appendix B and the comparative statics are derived below. □

The expected limit order surplus of the slow buyer increases in the arrival rate of the seller as it increases the value of the limit order and decreases the probability of trading with the market maker at the same time, provided that the value of the limit order is sufficiently high. Indeed, one has

$$\frac{\partial}{\partial\lambda_S}\left(V_{B,LO}\mathbb{P}[a^*-a'>v_\Delta\vee v_\Delta>a^*]\right)=\underbrace{\frac{\partial V_{B,LO}}{\partial\lambda_S}}_{>0}\underbrace{\mathbb{P}[\cdot]}_{>0}+\underbrace{\frac{\partial\mathbb{P}[\cdot]}{\partial\lambda_S}}_{>0}V_{B,LO}. \tag{5.50}$$

$\frac{\partial V_{B,LO}}{\partial\lambda_S}$ equals $e^{-\alpha\pi_B}\frac{\lambda_S}{\lambda_S-\frac{\alpha^2\sigma^2}{2}}\left(\frac{\lambda_S}{\lambda_S-\frac{\alpha^2\sigma^2}{2}}-1\right)$, which is positive since $\lambda_S-\frac{\alpha^2\sigma^2}{2}>0$. The trade probability is decreasing in λ_S by Equation (5.44), so the no-trade probability $\mathbb{P}[a^*-a'>v_\Delta\vee v_\Delta>a^*]$ increases in λ_S. The term $V_{B,LO}$ may in general take negative values. It will remain positive if $e^{\alpha\pi_B}>\frac{\lambda_S}{\lambda_S-\frac{\alpha^2\sigma^2}{2}}$, which holds for a broad range of parameters. Thus the buyer's expected limit order surplus is typically increasing in λ_S.

The expected limit order surplus of the slow buyer, $V_{B,LO}\mathbb{P}[a^*-a' > v_\Delta \vee v_\Delta > a^*]$, increases in latency Δ as a whole, because the probability of slow buyer's arrival increases. As in the case of the probability of trading, we are interested in the effects per unit of time. As shown above, because the dispersion of the fundamental value rises with Δ, the probability per unit of time Δ that the fundamental value will fall within the execution range, falls. Thus, the time-normalized effect of an increase in Δ on the expected limit order profit of the slow buyer is positive if the outside utility of the slow buyer is positive, $V_{B,LO} > 0$. Formally,

$$\frac{\partial}{\partial \Delta}\left(V_{B,LO}\frac{\mathbb{P}[a^*-a' > v_\Delta \vee v_\Delta > a^*]}{\Delta}\right) = \log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)\left(1 - \frac{\lambda_S e^{-\alpha\pi_B}}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)\frac{\lambda}{4\sqrt{3}\Delta^{\frac{3}{2}}\sigma\alpha} > 0. \tag{5.51}$$

Furthermore, the slow buyer's expected profit increases in λ as trading between the buyer and the market maker becomes more likely. Formally,

$$\frac{\partial}{\partial \lambda}\left(V_{B,LO}\frac{\mathbb{P}[a^*-a' > v_\Delta \vee v_\Delta > a^*]}{\Delta}\right) = \left(1 - \frac{\lambda_S e^{-\alpha\pi_B}}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)\left(1 - \frac{\log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)}{2\sqrt{3}\sqrt{\Delta}\sigma\alpha}\right) > 0. \tag{5.52}$$

Let us now consider the sensitivity of the market order surplus before analyzing the overall welfare of the slow buyer. The expected market order surplus of the slow buyer is decreasing in λ_S if the outside utility of the slow buyer is positive, $V_{B,LO} > 0$.

$$\frac{\partial}{\partial \lambda_S}\mathbb{E}\left[\left(1 - e^{-\alpha(\pi_B + v_\Delta - a^*)}\right)\mathbb{1}_{a^*-a' \le v_\Delta \le a^*}\right] = \frac{\alpha\sigma\sqrt{\Delta}e^{-\alpha\pi_B}\left(\alpha^2\sigma^2 e^{\alpha\pi_B} - 2\lambda_S\left(e^{\alpha\pi_B} - 1\right)\right)}{2\sqrt{3}\lambda_S\left(\alpha^2\sigma^2 - 2\lambda_S\right)^2} \tag{5.53}$$

This expression is negative provided that

$$1 - \frac{e^{-\alpha\pi_B}\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}} > 0 \tag{5.54}$$

i.e. if

$$V_{B,LO} > 0. \tag{5.55}$$

To analyze the sensitivity of the market order surplus with respect to Δ, we use again the profit per unit of time by dividing the expression by Δ. One has

$$\frac{\partial}{\partial \Delta}\left(\frac{\mathbb{E}\left[\left(1 - e^{-\alpha(\pi_B + v_\Delta - a^*)}\right)\mathbb{1}_{a^*-a' \le v_\Delta \le a^*}\right]}{\Delta}\right) = -\frac{e^{-\alpha\pi_B\left(e^{\alpha\pi_B}\log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right) - \frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}} + 1\right)}}{4\sqrt{3}\alpha\sqrt{\Delta^3}\sigma}. \tag{5.56}$$

This expression is negative provided that

$$e^{\alpha \pi_B} > \frac{\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}} - 1}{\log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)} = \frac{\frac{\partial a'}{\partial \sigma^2}}{\frac{a'}{\sigma^2}}. \tag{5.57}$$

The expected market order surplus of the slow buyer is decreasing in Δ if $e^{\alpha \pi_B} > \frac{\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}} - 1}{\log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)}$. In words, the expected payoff to a market order will decrease in latency if the private valuation is high enough compared to the elasticity of the reservation ask price with respect to the variance of the fundamental value σ^2.

Turning now to the impact of the model parameters on the slow buyer's overall welfare $V_{B,LO}\mathbb{P}[a^* - a' > v_\Delta \vee v_\Delta > a^*] + V_{B,MO}$, the effect of the seller's arrival rate can be computed as

$$\frac{\partial}{\partial \lambda_S}\left(V_{B,LO}\mathbb{P}[a^* - a' > v_\Delta \vee v_\Delta > a^*] + V_{B,MO}\right) = \frac{\alpha\sigma\sqrt{\Delta}\lambda e^{-\alpha\pi_B}\left(2\sqrt{3}\alpha\sqrt{\Delta}\sigma - \log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)\right)}{\sqrt{3}\left(\lambda_S - \frac{\alpha^2\sigma^2}{2}\right)^2}. \tag{5.58}$$

This expression is positive if

$$2\sqrt{3}\sigma\sqrt{\Delta} > a'. \tag{5.59}$$

Thus, the buyer's overall utility increases in λ_S if the reservation ask price is lower than the dispersion of the fundamental value $2\sqrt{3}\sigma\sqrt{\Delta} > a'$, which can be interpreted as a feasibility condition.

The overall expected payoff of the slow buyer also increases in his arrival rate λ as both his expected payoff due to market and limit order increase in λ.

Turning now to the effect of latency Δ, we are again interested in welfare per unit of time. One has:

$$\frac{\partial}{\partial \Delta}\left(\frac{V_{B,LO}\mathbb{P}[a^* - a' > v_\Delta \vee v_\Delta > a^*] + V_{B,MO}}{\Delta}\right) = \frac{\lambda e^{-\alpha\pi_B}\left(\frac{\alpha^2\sigma^2}{2} - \lambda_S \log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)\right)}{2\sqrt{3}\sqrt{\Delta^3}\alpha\sigma\left(\lambda_S - \frac{\alpha^2\sigma^2}{2}\right)}. \tag{5.60}$$

This expression is negative if

$$\frac{\alpha^2\sigma^2}{2} < \lambda_S \alpha a'. \tag{5.61}$$

The overall expected payoff of the slow buyer decreases in latency if the risk-aversion adjustment is lower than a term proportional to his reservation ask price $\frac{\alpha^2\sigma^2}{2} < \lambda_S \alpha a'$. Inserting the value of a' yields the condition

$$\frac{\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}} - 1}{\log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)} < \frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}. \tag{5.62}$$

Setting $y = \frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}$, Condition (5.62) can be written as $y - 1 - y\log y < 0$. Because $\lambda_S - \frac{\alpha^2\sigma^2}{2} > 0$, y varies between 1 and ∞. At $y = 1$, the above expression equals zero. Futhermore, the derivative of the expression is equal to $-\log y$, which is negative for $y > 1$. Taken together this means that the slow buyer's overall welfare is decreasing in latency.

The effects of latency, order size, and volatility are illustrated in Figure 5.7. The left panel shows the negative impact of latency on welfare for different order sizes Q. The middle panel shows the impact of latency and volatility, and the right panel the impact of order size and volatility, where we illustrate the effect of Equation (5.59) by choosing $\Lambda_S = 10$ for the top panels and $\Lambda_S = 100$ for the bottom panels.

5.7. Conclusion

Are faster markets better for institutional investors? The current paper presents a model considering the impact of both order size and latency on the price impact of trades. A model is proposed, which is solved analytically and is novel in the theoretical microstructure literature. Larger latency increases adverse selection costs to the market maker and reduces his probability of trading with a slow investor. A larger order size decreases the slow trader's outside option, making him susceptible to accept a worse price for his trade. It is shown that the first order effect of increased latency and increased order size is to increase price impact. Their joint impact is also positive. When the probability of trading is taken into consideration, the utility of the slow institutional investor decreases with latency. Furthermore, this model is surmisable to possible extensions. Natural extensions of this work include taking the order size as an endogenous variable in a dynamic model, creating scope for trading influencing prices every round and the market maker learning from such price changes. It would also be interesting to observe how such a model could be extended to include the market maker's inventory management or competition among different high-frequency traders that are possibly subject to different latencies.

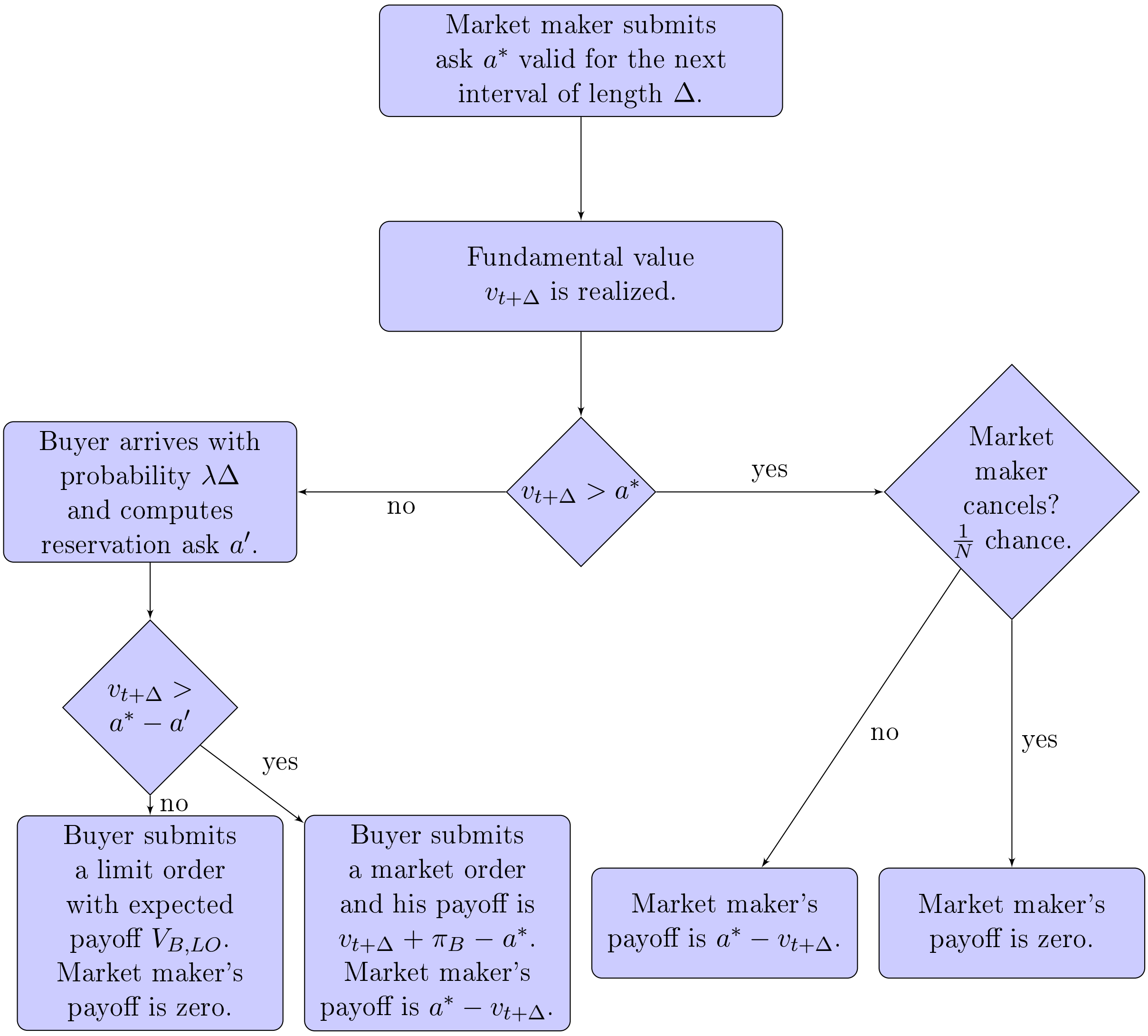

Figure 5.1: Flow of events.

This figure shows the flow of events. For simplicity, the time t fundamental value is normalized to zero, $v_t = 0$.

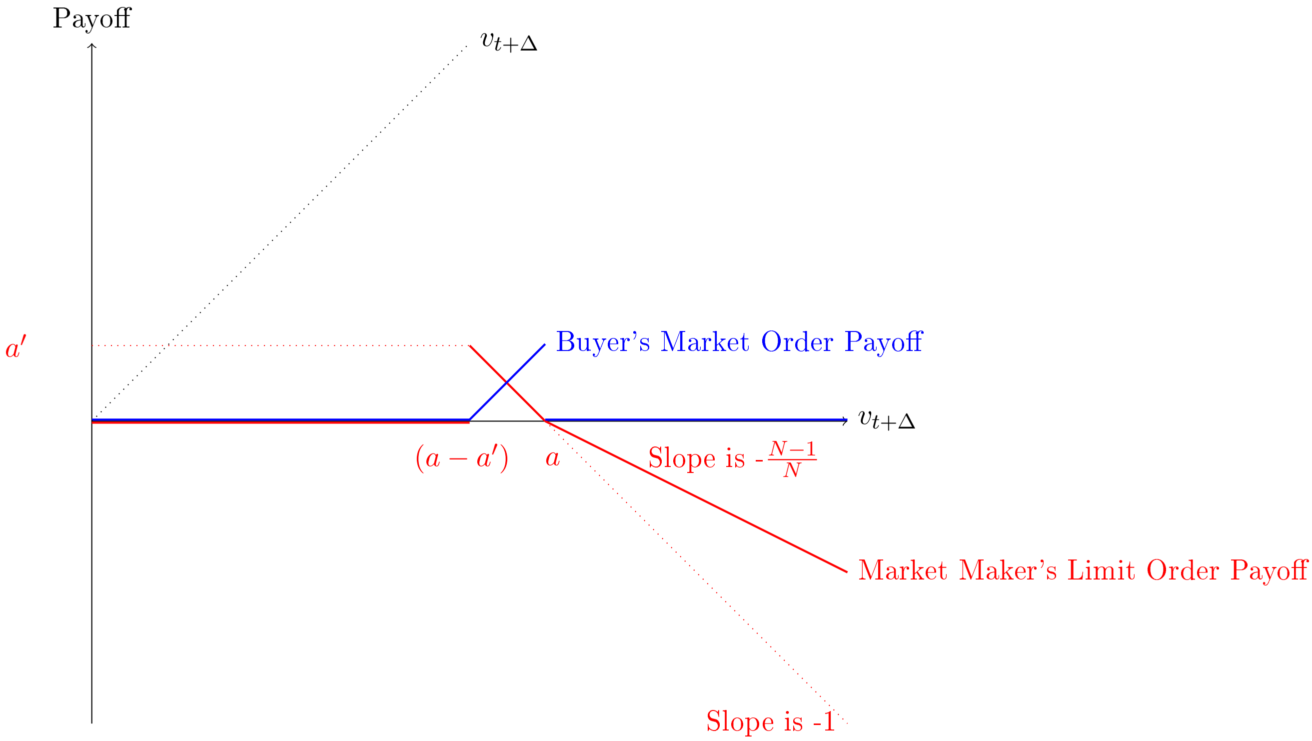

Figure 5.2: Market maker's and buyer's payoff structure as a function of the fundamental value $v_{t+\Delta}$. This figure shows market maker's and buyer's payoff structure as a function of the fundamental value $v_{t+\Delta}$. There are N high-frequency traders in the market. Market maker submits ask at price a and the maximum price at which the buyer is willing to buy is the reservation ask, a'.

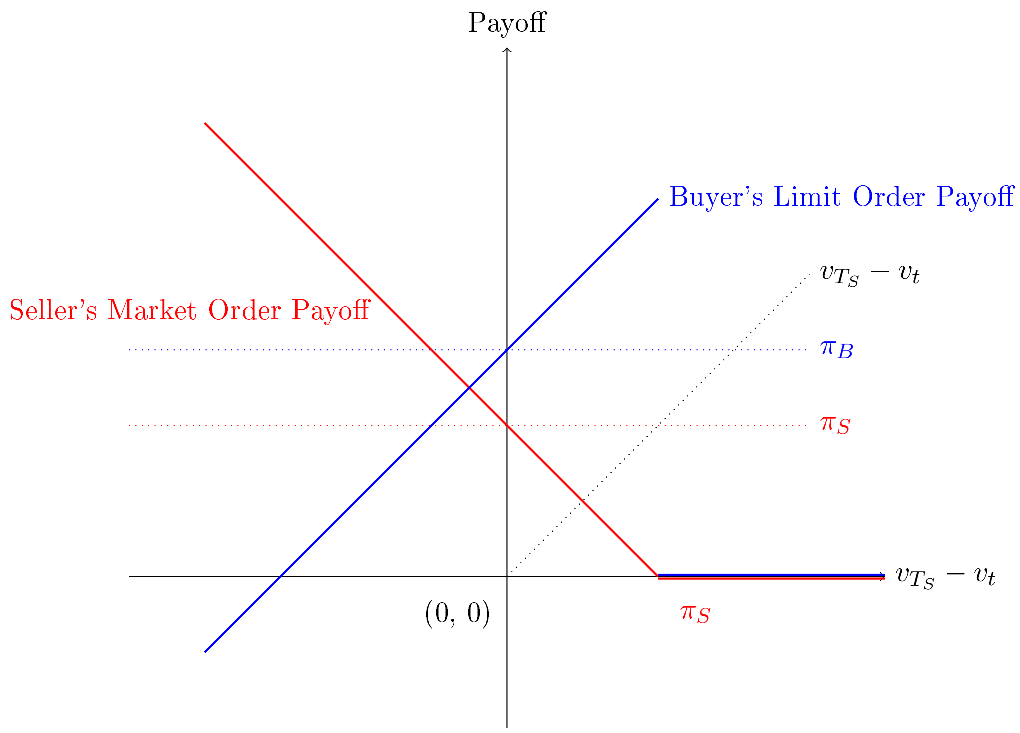

Figure 5.3: Buyer's and seller's payoff structure as a function of the change in the fundamental value $v_{T_S} - v_t$. This figure shows buyer's and seller's payoff structure as a function of the change in the fundamental value $v_{T_S} - v_t$. The buyer submits a bid at price v_t and the minimum price at which the seller is willing to sell is his private valuation π_S.

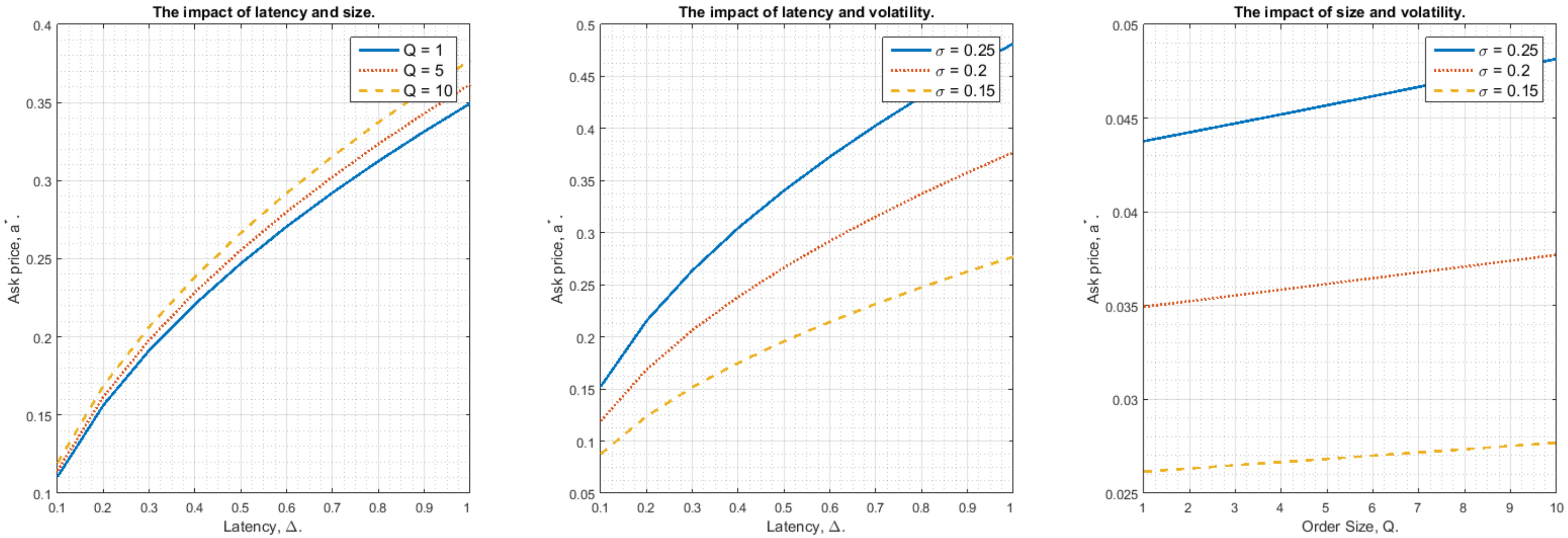

Figure 5.4: Market maker's ask price as a function of latency Δ and order size Q.
This figure illustrates the sensitivity of the ask price a^* to the latency Δ, order size Q, and volatility σ. The base parameters chosen are $\sigma = 0.2$, $\lambda = 1$, $\Lambda_S = 10$, and $Q = 1$. The left panel shows the impact of latency and order size. The middle panel shows the impact of latency and volatility, and the right panel the impact of order size and volatility.

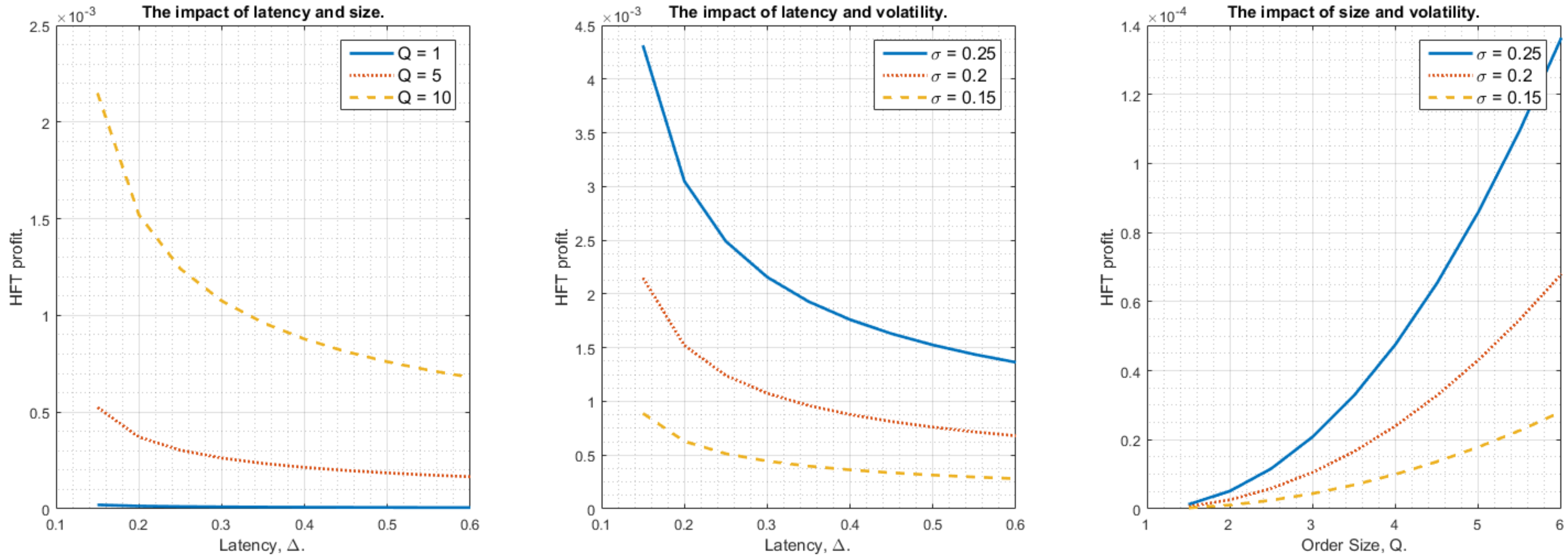

Figure 5.5: Expected profit of the market maker as a function of latency Δ and order size Q. This figure illustrates the sensitivity of the market maker's profit to the latency Δ, order size Q, and volatility σ. The base parameters chosen are $\sigma = 0.2$, $\Lambda_S = 10$, and $Q = 1$. The left panel shows the impact of latency and order size. The middle panel shows the impact of latency and volatility, and the right panel the impact of order size and volatility.

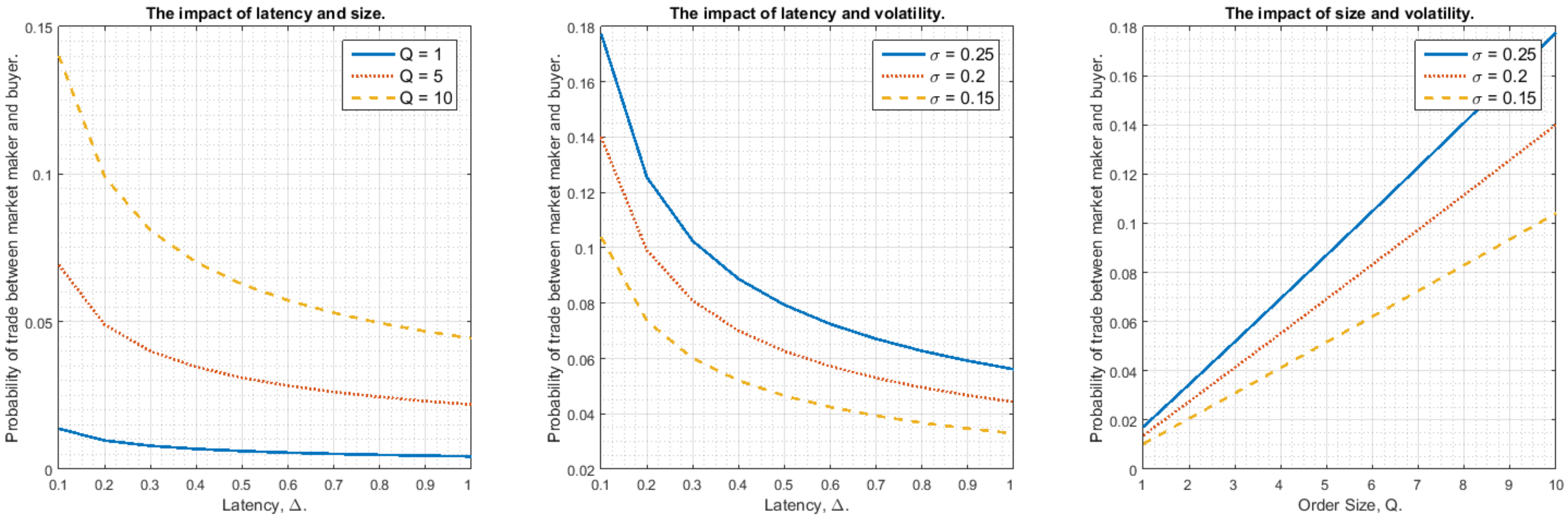

Figure 5.6: Probability of trade between market maker and buyer as a function of latency Δ and order size Q. This figure illustrates the sensitivity of the probability of trading to the latency Δ, order size Q, and volatility σ. The base parameters chosen are $\sigma = 0.2$, $\Lambda_S = 10$, $Q = 1$, and $\lambda = 1$. The left panel shows the impact of latency and order size. The middle panel shows the impact of latency and volatility, and the right panel the impact of order size and volatility.

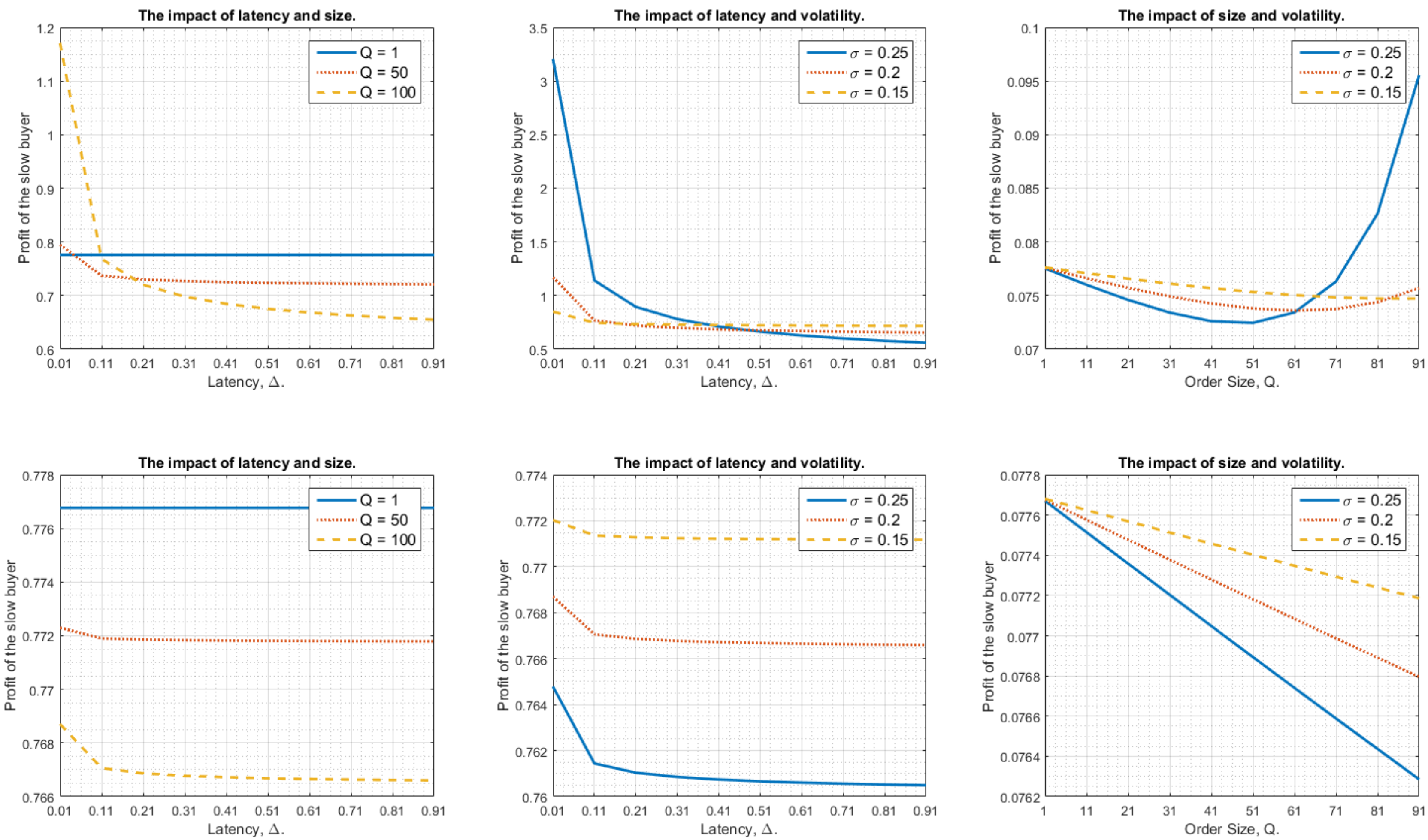

Figure 5.7: Expected profit of the slow buyer as a function of latency Δ and order size Q.
This figure illustrates the sensitivity of the slow buyer's profit to latency Δ, order size Q, and volatility σ. The base parameters chosen are $\sigma = 0.2$, $\Lambda_S = 10$, and $Q = 1$. The left panel shows the impact of latency and order size. The middle panel shows the impact of latency and volatility, and the right panel the impact of order size and volatility. The top three panels are computed with $\Lambda_S = 10$, and the bottom three panels with $\Lambda_S = 100$.

Part III

Appendices

Chapter 6

Formal Description of the Model in Chapter 3

This Appendix provides a detailed formal description of the model. Our description follows that provided in Goettler et al. (2009).

Structure of the Trading Game

We define an action a trader takes as $a = (p, q, x)$, where p is the price at which she submits an order, $q \geq 0$ the priority of her order and x is $+1$ or -1 when the order is buy or sell and 0 if there is no order. The priority of an order is determined by p, x and the current book L. q decreases with increasing priority in the queue. If the order is a market order, the priority is 0. The priority of a limit order is the new order's position in the queue. Formally:

$$q(p, x) = \begin{cases} 0 & \text{if } x = 0 \text{ or} \\ & x = 1,\, p \geq A(L) \text{ or} \\ & x = -1,\, p \leq B(L), \\ |l^p + x| & \text{otherwise.} \end{cases}$$

A trader reentering the market at time t who does not have a pending order in the book faces the same problem as a new trader. A trader having an order in the book additionally has the option of leaving the order unchanged, potentially taking advantage of an improvement in the priority of the order. The optimal action the trader chooses depends on the state in which the trader enters the market.

Let $s(\theta)$ be the state observed on a particular entry to the market by a trader of type θ. The state $s(\cdot)$ includes:

1. the history of the game and its elements as described above, the changes of the fundamental value until time $t - \Delta$ if $h = 0$ or until time t if $h = 1$, where Δ is the lag (if any) with which slow traders observe the fundamental value. The state further contains the limit order book and the status of the previous action $a = (p, q, x)$ if the trader previously submitted a limit order,

where p is the price at which the order was submitted, q the current priority at price p and x an indicator of a buy ($+1$) or sell (-1) order or no order (0). If the trader is entering for the first time, x is set to zero;

2. a variable $z \in \{0, 1\}$ used in the Bellman equation to set the agent's future payoff to zero once she trades. The variable z can also be viewed as the trader's budget or the number of shares she has available to trade. In our model, traders enter the market with $z = 1$.

For our numerical implementation, we have to limit the possibly infinite state space. First, because the fundamental value follows a random walk, it can take unbounded values. For this reason, we only track all prices and quotes relative to the fundamental value. We further collapse the characterization of market conditions into (i) the fundamental value v_t for $h = 1$ and $v_{t-\Delta}$ if $h = 0$; (ii) the price of the last transaction p_t and whether the transaction was buyer- or seller-initiated $b_t = -1$ or $b_t = 1$; and (iii) the limit order book, which is represented by bid and ask prices B_t, A_t, bid and ask sizes l_t^B, l_t^A, cumulative depths at buy and sell $D_t^b = \sum_{i=0}^{N} (l_t^i > 0), D_t^s = \sum_{i=0}^{N} (l_t^i < 0)$. We investigated models with broader market characteristics and found that while convergence occurs more slowly, the results remain practically unchanged.

Traders can choose an action from a feasible action set. In order not to hamper computational tractability, we restrict the number of ticks where traders can place limit orders to k. We choose k such that it does not influence equilibrium strategies. Denote the expectation of the fundamental value given the state s as $\hat{v}(s) = \mathbb{E}(v|s)$. The feasible action set is then formally defined as

$$\mathcal{A}(s) = \{(p, q, x) \mid (i)\, x \in \{-1, 0, 1\},\ (ii)\, q = \hat{q}(p, x),\ (iii) \text{ if } q \neq 0 \Rightarrow p \in [\hat{v}(s) - k, \hat{v}(s) + k] \cap \mathcal{P}\}.$$

A mixed strategy for a trader of type θ is $\sigma_\theta : S_\theta \longrightarrow \prod_{s \in S_\theta} \Delta(\mathcal{A}(s))$, where S_θ is the set of feasible states that a trader of type θ may encounter and $\Delta(\mathcal{A}(s))$ represents possible probability distributions over $\mathcal{A}(s)$.

Every action leads to an expected payoff which is composed of two parts. The first is the payoff from execution prior to reentry and the second the continuation value if the trader reenters prior to the execution of her order. We consider symmetric equilibria and denote by $\sigma = \{\sigma_\theta\}_{\theta \in \Theta}$ the strategies adopted by every other player. Let $\phi(\tau, v; s, \tilde{a}, \sigma)$ be the probability that an action $\tilde{a} = (\tilde{p}, \tilde{q}, \tilde{x})$ taken at time s leads to an execution at time τ when the fundamental value is v and all other players' strategies are σ. We further let $f(v|s,t)$ denote the density function of the fundamental value at time t given state s. Suppose the trader reenters the market at time $w > 0$. Her expected payoff due to execution prior to reentry is

$$\pi(s, \tilde{a}, w, \sigma) = \int_0^w \int_{-\infty}^{\infty} e^{-\rho t} \tilde{x}(\alpha + v_t - \tilde{p})\phi(\cdot) f(v|s,t) dv dt,$$

where the inner integral is over the possible values of the fundamental value and the outer integral is over the possible times at which execution can occur. Inside the integral is the instantaneous payoff discounted back to the time at which the order was submitted.

The agent chooses her action such that it maximizes her expected payoff (the sum of the payoff due to execution prior to reentry and the continuation value). The reentry time is distributed randomly and exogenously according to $G(\cdot)$. The state s' in which the trader reenters has probability of occurring $v(s'|s, \tilde{a}, w, \sigma)$. We denote the trader's value function in state s by $J(s)$. The trader's optimization problem can be written as a dynamic programming problem, where the Bellman equation is

$$J(s, \sigma) = \max_{\tilde{a} \in \mathcal{A}(s)} \int_0^{\infty} \left(\pi(s, \tilde{a}, w, \sigma) + e^{-\rho t} \int_{s' \in S_\theta} J(s', \sigma) v(s'|s, \tilde{a}, w, \sigma) ds' \right) dG(w).$$

Since the trader always faces a maximization problem over a well-defined and finite action set, the maximum over all feasible actions exists.

Fixing the strategies of all other traders, a pure strategy y^*_θ for a trader of type θ is a best response if and only if for every $s \in S_\theta$

$$y^*_\theta(s) = \arg \max_{\tilde{a} \in \mathcal{A}(s)} \int_0^{\infty} \left(\pi(s, \tilde{a}, w, \sigma) + e^{-\rho t} \int_{s' \in S_\theta} J(s', \sigma) v(s'|s, \tilde{a}, w, \sigma) ds' \right) dG(w).$$

A collection of strategies $y^* = \{y^*_\theta\}_{\theta \in \Theta}$ is a Markov perfect equilibrium if and only if for each pair $\theta \in \Theta, y^*_\theta$ is a best response in every feasible state $s \in S_\theta$.

Beliefs and Q-Learning

Equilibrium is obtained by finding common beliefs $Q(a|s)$, where s is the current state and $a = a(s)$ an attainable action at this state. This is done by simulating the market and updating beliefs until they converge. In finding the fixed-point of this game, since an analytic solution is not tractable, we turn to the stochastic algorithm of Pakes and McGuire (2001) implemented for a similar kind of games in Goettler et al. (2005, 2009).

One way to obtain beliefs would be to update them by integrating over all possible sequences of future outcomes that lead to a transaction being executed. By contrast, the algorithm of Pakes and McGuire (2001) tracks each share in the book until it executes in the market simulation. Upon execution, we update the belief $Q(a|s)$ at the state from which the order was submitted by averaging this outcome with the previous outcomes for shares submitted at this state discounted back by the time it took the share to execute. We use an online implementation of this reinforcement or Q-learning algorithm and update the $Q(a|s)$ with the new $Q(a'|s')$ from the new state s' where the trader's action was a', not just relying on the final executions, but incorporating the new belief as soon as a trader reenters and chooses a new action. Figure 3.8 provides a diagrammatic representation of this procedure.

Not every possible state is visited during the simulation, thus it focuses only on the recurrent subset of the state space.[1] The advantage of this approach is that updates are computed only for the states actually visited, thus lowering the memory and computation time requirements. The disadvantage is that not all relevant states might have been visited during the simulation. To prevent the algorithm from falling into local equilibria which do not satisfy the perfection condition, a couple of state space exploration techniques are implemented, namely: (i) optimistic initial beliefs – when beliefs of the first explored state-action get corrected, the other un-explored state still possesses higher optimistic initial beliefs and is thus likely to be explored, (ii) trembles – in the learning phase, traders choose suboptimal actions with a small probability, and (iii) resetting the learning speed from time to time in order to speed up learning in different phases of the simulation.

The simulation has two main phases: (i) the learning phase, where the beliefs are updating and the exploration of the state space is promoted, and (ii) the simulation of the equilibrium, during which the beliefs are fixed and the exploration is not artificially promoted. The algorithm switches between phase (i) and phase (ii) depending on the convergence of the beliefs. The convergence during the first phase is called convergence of the first type or learning convergence and if beliefs are found to have converged in the first phase, the second phase with fixed beliefs starts. Convergence of the beliefs in the second phase is called second type convergence or equilibrium convergence. If second type convergence is not

[1] A recurrent subset of states has the following properties: (i) regardless of the initial state, the system eventually enters the recurrent class, (ii) once entered, the probability of each state outside the recurrent class is zero, and (iii) each state in the recurrent class is visited infinitely often as t approaches infinity.

satisfied, the algorithm returns to the learning phase. When second type convergence is satisfied, we claim that the fixed point was obtained and the data collected during the simulation phase represent equilibrium values.

Formally, at each time t in each state s encountered in the simulation, each action $\tilde{a}$ has an associated payoff $Q_t(\tilde{a}|s)$. This real number represents the current belief of an agent about the payoff from this action at this state.[2] The beliefs at each time t imply an optimal strategy y_t, which assigns the payoff maximizing action at each state, $a^*(s) \in \arg\max_{\tilde{a}\in\mathcal{A}(s)} Q_t(\tilde{a}|s)$. The value of state s then is $J(s, y_t) = Q_t(a^*(s)|s)$.

The value for a newly encountered state is called an initial belief $Q_0(\tilde{a}|s)$ and is determined in the following way. Consider a buy limit order. The initial belief is determined by the payoff $\alpha + \hat{v} - p$ discounted by the expected time until the arrival of a new trader for whom being a counterparty yields a non-negative payoff. Here, $\hat{v}(s) = \mathbb{E}(v|s)$ is the expectation of the current fundamental value given the trader's type. If the trader is a HFT, she knows the current fundamental value v_t. For the uninformed slow trader, we have to compute the initial expectation of the fundamental value. Let $\delta(s(\theta)) = \mathbb{E}(v|s) - v_{t-\Delta}$. We update this expected difference between the expected current fundamental value and the lagged fundamental value using the following updating rule:

$$\delta_{r+1}(s(\theta)) = \frac{r}{r+1}\delta_r(s(\theta)) + \frac{1}{r+1}(v_t - v_{t-\Delta}).$$

Thus, for an uninformed trader, the estimate of the current fundamental value is given by $\hat{v}(s(\theta)) = v_{t-\Delta} + \delta_r(s(\theta))$. The initial belief for market orders is straightforward. The initial belief of not submitting any order is determined as the average of all non-negative beliefs the trader has available at the current state discounted by the expected return time of the particular trader.

If the previous state was s and the state s' is hit, the continuation value $J(s'|y_{t'})$ is defined as in Figure 3.8 based on the action a' taken in the new state s':

1. *Market order*: payoff from the market order.
2. *Limit order*: expected value of the limit order represented by the action a', $Q(a'|s')$.
3. *No order*: expected value of taking no action a', $Q(a'|s')$.

[2] Q-learning was first described as a model for animals' learning in Watkins (1989) and more details and the underlying theory can be found in Bertsekas and Tsitsiklis (1996).

Put together, the Q-factors are updated in the following fashion, where $J(s'|y_{t'})$ represents the payoff to the action taken at state s' and time t'.

$$Q_{t'}(a^*|s) = \frac{n}{n+1} Q_t(a^*|s) + \frac{1}{n+1} e^{-\rho(t'-t)} J(s'|y_{t'})$$

Here, $n(\tilde{a}^*, s)$ is a positive integer that is incremented each time $\tilde{a}^*$ is chosen in state s. Periodically during the simulation, we restart n to some small initial value of n_0 for some action and state pairs to obtain quicker convergence.

Similarly, if the previously submitted limit order is executed, the expected payoff at the previous state is updated by $\tilde{x}(\alpha + v_{t'} - \tilde{a}^*)$, where t' is the current time.

A formal proof of convergence to the optimal Q-factors with probability 1 is provided in Section 5.6 of Bertsekas and Tsitsiklis (1996).

Convergence Criteria

We simulate a couple of billion events during which we continuously decrease the artificial exploration of the state space by decreasing the trembling probability in order to achieve a soft landing to the equilibrium values of the Q-factors.[3] We use the same convergence criteria as Goettler et al. (2009). After an initial exploration of the state space, we perform the following computations every 300 million events:

1. If $\frac{|Q_{t_2}^{k_2}(\tilde{a}|s) - Q_{t_1}^{k_1}(\tilde{a}|s)|}{k_2 - k_1}$ is small (less than 0.01), then
2. fix beliefs $Q(\cdot)$ and simulate 300 million events,
3. compare fixed beliefs to (a) "one step ahead" $\tilde{J}_1$ and (b) "realized" $\tilde{J}$ empirical payoffs. If
    - the correlation between J^* and both $\tilde{J}_1$ and $\tilde{J}$ exceeds 0.99, and
    - the mean absolute error in beliefs over n is less than 0.01,

   convergence is achieved.

[3]Increasing the trembling probability involves a tradeoff. The benefit is that it allows a faster exploration of the state space. The cost is that it also impacts the strategies of traders that do not tremble by making them consider a certain fraction of "erroneous orders" to be part of general market conditions.

Step (1) corresponds to the learning convergence criterion described above and step (3) corresponds to the equilibrium convergence criteria. If (1) is not satisfied, learning continues, and if (3) is not satisfied, the algorithm returns to the learning phase. Here, k_1 is the number of times that action $\tilde{a}$ has been chosen in state s at the start of the current 300 million events, and k_2 the number of times it has been chosen at the end of the current 300 million new events. Further, t_1 and t_2 represent the corresponding times.

The one step ahead and realized empirical payoffs required for these convergence computations are determined as follows. Note that eventually every trader in this model executes and leaves the market. At the time she executes, she obtains a "realized payoff", computed as follows. Suppose the trader enters at t and executes at t' . If $\tilde{a}$ was her most recent action before execution, her realized payoff is then $\tilde{J}(s, y^*) = e^{-\rho(t'-t)}\tilde{x}(\alpha + v_{t'} - \tilde{p})$. The "one step ahead" payoff is based on the trader's next entry time or execution time, whichever is sooner. Suppose a trader takes an action $\tilde{a}$ at time t, and reenters at $t' > t$ with a new state s'. Her one-step ahead empirical payoff is $\tilde{J}_1(s, y^*) = e^{-\rho(t'-t)} J^*(s', y^*)$. If the trader's order executes prior to reentry, her one-step ahead empirical payoff is $\tilde{J}_1(s, y^*) = e^{-\rho(t'-t)}\tilde{x}(\alpha + v_{t'} - \tilde{p})$.

Chapter 7

Proofs of Statements in Chapter 5

Appendix A. Price Impact as a Function of Latency

The following proves lemma 5.4.1 by computing the buyer's reservation utility.

Proof. The buyer's payoff is depicted in Figure 5.3. The fundamental value, v_{T_S}, is normally distributed with mean zero and variance $\sigma^2 T_S$. The seller, who comes at time T_S will only execute if $v_{T_S} \leq \pi_S$. Put altogether, the buyer's utility from submitting a limit order is given by the expectation of the exponential of a truncated normal variable weighted by the random arrival time T_S, which is exponentially distributed with intensity λ_S per unit of time. Formally, it is given by the following double integral

$$V_{B,LO} = 1 - \int_0^\infty \lambda_S e^{-\lambda_S y} \int_{-\infty}^{\pi_S} \frac{e^{-\alpha(\pi_B+x)}}{\sqrt{2\pi\sigma^2 y}} e^{-\frac{1}{2}\frac{x^2}{\sigma^2 y}} \mathrm{d}x\mathrm{d}y \tag{7.1}$$

$$= 1 - \lambda_S e^{-\alpha\pi_B} \int_0^\infty e^{-\lambda_S y + \frac{\alpha^2\sigma^2}{2}y} \int_{-\infty}^{\pi_S} \frac{1}{\sqrt{2\pi\sigma^2 y}} e^{-\frac{1}{2}\frac{(x+\alpha\sigma^2 y)^2}{\sigma^2 y}} \mathrm{d}x\mathrm{d}y \tag{7.2}$$

$$= 1 - \lambda_S e^{-\alpha\pi_B} \int_0^\infty e^{-\lambda_S y + \frac{\alpha^2\sigma^2}{2}y} \Phi(\frac{\pi_S + \alpha\sigma^2 y}{\sigma\sqrt{y}}) \mathrm{d}y, \tag{7.3}$$

where $\Phi(\cdot)$ represents the normal CDF. The second equation was obtained by expanding the x^2 by $\alpha\sigma^2 y$. The third equality was obtained by the change of variables $z = x + \alpha\sigma^2 y$. Let's focus on the integral term, which we denote I and solve by parts.

$$I := \int_0^\infty e^{-\lambda_S y + \frac{\alpha^2\sigma^2}{2}y} \Phi(\frac{\pi_S + \alpha\sigma^2 y}{\sigma\sqrt{y}}) \mathrm{d}y \tag{7.4}$$

$$= -\frac{1}{\lambda_S - \frac{\alpha^2\sigma^2}{2}} \left[e^{-y(\lambda_S - \frac{\alpha^2\sigma^2}{2})} \Phi(\frac{\pi_S + \alpha\sigma^2 y}{\sigma\sqrt{y}}) \right]_0^\infty - \frac{(-1)1}{\lambda_S - \frac{\alpha^2\sigma^2}{2}} \int_0^\infty e^{-\lambda_S y + \frac{\alpha^2\sigma^2}{2}y} \frac{\partial}{\partial y} \Phi(\frac{\pi_S + \alpha\sigma^2 y}{\sigma\sqrt{y}}) \mathrm{d}y \tag{7.5}$$

$$= \frac{1}{\lambda_S - \frac{\alpha^2\sigma^2}{2}} + \frac{1}{\lambda_S - \frac{\alpha^2\sigma^2}{2}} \int_0^\infty e^{-y(\lambda_S - \frac{\alpha^2\sigma^2}{2}) - \frac{1}{2}(\frac{\pi_S + \alpha\sigma^2 y}{\sigma\sqrt{y}})^2} \left(\frac{\alpha\sigma}{2\sqrt{y}} - \frac{\pi_S}{2\sqrt{y^3}\sigma} \right) \mathrm{d}y \tag{7.6}$$

$$= \frac{1}{\lambda_S - \frac{\alpha^2\sigma^2}{2}} \left(1 + e^{-\alpha\pi_S} \int_0^\infty e^{-y\lambda_S - \frac{\pi_S}{2\sigma^2 y}} \left(\frac{\alpha\sigma}{2\sqrt{y}} - \frac{\pi_S}{2\sqrt{y^3}\sigma} \right) \mathrm{d}y \right). \tag{7.7}$$

Here, the second equality comes from solving the I integral by parts. Computing the limits of the left term and the partial derivative for the right term yields the third equality. The fourth results from collecting terms. Let us again compute the integral part separately.

$$
\begin{aligned}
J &:= \int_0^\infty e^{-y\lambda_S - \frac{\pi_S}{2\sigma^2 y}} \left(\frac{\alpha\sigma}{2\sqrt{y}} - \frac{\pi_S}{2\sqrt{y^3}\sigma} \right) \mathrm{d}y && (7.8)\\
&= \int_0^\infty e^{-y\lambda_S - \frac{\pi_S}{2\sigma^2 y}} \left(\frac{\alpha\sigma}{2\sqrt{y}} \right) \mathrm{d}y - \int_0^\infty e^{-y\lambda_S - \frac{\pi_S}{2\sigma^2 y}} \left(\frac{\pi_S}{2\sqrt{y^3}\sigma} \right) \mathrm{d}y && (7.9)\\
&= \left(\frac{\alpha\sigma}{2}\right) \int_0^\infty \frac{1}{\sqrt{y}} e^{-\frac{1}{2}(2\lambda_S y + \frac{\pi_S}{\sigma^2 y})} \mathrm{d}y - \left(\frac{\pi_S}{2\sigma}\right) \int_0^\infty \frac{1}{\sqrt{y^3}} e^{-\frac{1}{2}(\lambda_S y + \frac{\pi_S}{\sigma^2 y})} \mathrm{d}y && (7.10)\\
&= J_1 - J_2. && (7.11)
\end{aligned}
$$

The above separation is useful, because the two integrals are special Bessel functions found in Lebedev (1972) Sects. 8.432 6 p. 959, and 8.469 3 p. 967:

$$
\begin{aligned}
\int_0^\infty \frac{1}{\sqrt{t}} e^{-\frac{\delta}{2}(t+\frac{1}{t})} \mathrm{d}t &= 2\sqrt{\frac{\pi}{2\delta}} e^{-\delta}, && (7.12)\\
\int_0^\infty \frac{1}{\sqrt{t^3}} e^{-\frac{1}{2}(\beta t + \frac{\gamma}{t})} \mathrm{d}t &= \sqrt{\frac{2\pi}{\gamma}} e^{-\sqrt{\beta\gamma}}, && (7.13)\\
& && (7.14)
\end{aligned}
$$

where $\beta = 2\lambda_S$, $\gamma = \frac{\pi_S}{\sigma^2}$ and $\delta = \sqrt{\beta\gamma}$. Hence,

$$
\begin{aligned}
J_1 &= \frac{\alpha\sigma}{2} \sqrt{\frac{\pi}{\lambda_S}} e^{\sqrt{-2\lambda_S \frac{\pi_S}{\sigma^2}}}, && (7.15)\\
J_2 &= \sqrt{\frac{\pi}{2}} \sqrt{\pi_S} e^{-\sqrt{2\lambda_S \frac{\pi_S}{\sigma^2}}}. && (7.16)
\end{aligned}
$$

Inserting $J = J_1 - J_2$ back into the integral $I := \frac{1}{\lambda_S - \frac{\alpha^2\sigma^2}{2}} \left(1 + e^{-\alpha\pi_S} J\right)$ and I back into the utility of submitting a limit order, $V_{B,LO} = 1 - \lambda_S e^{-\alpha\pi_B} I$ yields the following

$$
V_{B,LO} = 1 - \frac{\lambda_S e^{-\alpha\pi_B}}{\lambda_S - \frac{\alpha^2\sigma^2}{2}} \left(1 + e^{-\alpha\pi_S} e^{-\sqrt{\frac{2\lambda_S\pi_S}{\sigma^2}}} \left[\frac{\alpha\sigma}{2} \sqrt{\frac{\pi}{\lambda_S}} - \sqrt{\frac{\pi}{2}} \sqrt{\pi_S} \right] \right), \quad (7.17)
$$

which concludes the proof of the buyer's reservation utility from submitting a limit order and waiting for a seller.

□

Appendix B. Welfare Analysis of Market Maker's Latency

Proof. The following proves the expected utility of the slow buyer from a market order, Equation (5.49):

$$
\begin{aligned}
\mathbb{E}\left[1 - e^{-\alpha(\pi_B + v_\Delta - a^*)} | a^* - a' \le v_\Delta \le a^*\right] &= \lambda\Delta \int_{a^*-a'}^{a^*} \frac{1 - e^{-\alpha(-a^* + \pi_B + x)}}{2\sqrt{3\Delta}\sigma} \mathrm{d}x && (7.18)\\
&= \lambda\Delta \frac{e^{-\alpha\pi_B\left(\alpha a' e^{\alpha\pi_B} - e^{\alpha a'} + 1\right)}}{2\sqrt{3}\alpha\sqrt{\Delta}\sigma} && (7.19)\\
&= \lambda\sqrt{\Delta} \frac{\log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right) + e^{-\alpha\pi_B}\left(1 - \frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)}{2\sqrt{3}\alpha\sigma}. && (7.20)
\end{aligned}
$$

The first and second equalities follow directly from computing the expectation. The third equality is obtained by plugging in the value for the reservation ask price a' from lemma 5.4.2. □

Proof. The following proves the expected utility of the slow buyer from a limit order, Equation (5.48). Let us first compute the probability of not executing upon slow buyer's arrival:

$$\mathbb{P}[a^* - a' > v_\Delta \vee v_\Delta > a^*] = \lambda\Delta \left(\int_{-\sqrt{3\Delta}\sigma}^{a^*-a'} \frac{1}{2\sqrt{3\Delta}\sigma} \, \mathrm{d}x + \int_{a}^{\sqrt{3\Delta}\sigma} \frac{1}{2\sqrt{3\Delta}\sigma} \, \mathrm{d}x \right) \tag{7.21}$$

$$= \lambda\Delta \left(1 - \frac{a'}{2\sqrt{3}\sqrt{\Delta}\sigma} \right) \tag{7.22}$$

$$= \lambda\Delta \left(1 - \frac{\log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)}{2\sqrt{3}\alpha\sqrt{\Delta}\sigma} \right). \tag{7.23}$$

The first and second equalities follow directly from computing the probability. The third equality is obtained by plugging in the value for the reservation ask price a' from lemma 5.4.2.

Multiplying this probability by the expected payoff from a limit order $V_{B,LO}$ given in lemma 5.4.1 yields

$$V_{B,LO}\mathbb{P}[a^* - a' > v_\Delta \vee v_\Delta > a^*] = \lambda\Delta \left(1 - \frac{\log\left(\frac{\lambda_S}{\lambda_S - \frac{\alpha^2\sigma^2}{2}}\right)}{2\sqrt{3}\alpha\sqrt{\Delta}\sigma} \right) \left(1 - \frac{\lambda_S e^{-\alpha\pi_B}}{\lambda_S - \frac{\alpha^2\sigma^2}{2}} \right). \tag{7.24}$$

□

Bibliography

[1] Aït-Sahalia, Y. and J. Jacod. 2014. *High-Frequency Financial Econometrics*. Princeton, New Jersey, USA: Princeton University Press.

[2] Aït-Sahalia, Y. and M. Sağlam. 2014. High frequency traders: Taking advantage of speed. *Working paper*, National Bureau of Economic Research.

[3] Baldauf, M. and J. Mollner. 2015. High-Frequency Trading and Market Performance. *Working Paper.*

[4] Baruch, S. and L. R. Glosten, 2013. Fleeting orders. *Working paper*, Columbia University and the University of Utah.

[5] Becchetti, L., Ferrari, M. and U. Trenta. 2014. The impact of the French Tobin tax, *Journal of Financial Stability 15*, 127–148.

[6] Bernales, A. 2014. How fast can you trade? High frequency trading in dynamic limit order markets. *Working paper*, Banque de France.

[7] Bessembinder, H. 2003. Issues in assessing trade execution costs. *Journal of Financial Markets 6*, 233–257.

[8] Biais, B., T. Foucault and S. Moinas. 2015. Equilibrium fast trading. *Journal of Financial Economics 116(2)*, 292–313.

[9] Bouchaud, J. P. 2010. Price impact. *Encyclopedia of Quantitative Finance, Wiley Online Library.*

[10] Brogaard, J. 2010. High Frequency Trading and Its Impact on Market Quality. *Working paper*, Northwestern University.

[11] Brogaard, J., T. J. Hendershott and R. Riordan. 2014. High-frequency trading and price discovery. *The Review of Financial Studies 27*, 2267–2306.

[12] Brogaard, J., A. Carrion, T. Moyaert, R. Riordan, A. Shkilko, and K. Sokolov (2015). High-frequency trading and extreme price movements. *University of Washington Working Paper.*

[13] Budish, E. B., P. Cramton and J. J. Shim. 2013. The high-frequency trading arms race: Frequent batch auctions as a market design response. *Fama-Miller Working Paper, 14–03.*

[14] Carrion, A. 2013. Very fast money: high-frequency trading on NASDAQ. *Journal of Financial Markets 16*, 680–711.

[15] Chacko, G. C., J. W. Jurek and E. Stafford. 2008. The price of immediacy. *Journal of Finance 63*, 1253–1290.

[16] Chakravarty, S., P. Jain, J. Upson, and R. Wood. 2012. Clean sweep: Informed trading through intermarket sweep orders. *Journal of Financial and Quantitative Analysis 47*, 415–435.

[17] Colliard, J.-E. and T. Foucault. 2012. Trading Fees and Efficiency in Limit Order Markets, *Review of Financial Studies 25*, 3389–3421.

[18] Colliard, J. E., and P. Hoffmann. 2015. Financial Transaction Taxes, Market Composition, and Liquidity. *Working paper*, European Central Bank.

[19] Collin-Dufresne, P. and V. Fos. 2015. Do prices reveal the presence of informed trading? *Journal of Finance 70*, 1555–1582.

[20] Committee of European Securities Regulators. 2010. Micro-structural issues of the European equity markets; Ref.: CESR/10-142, *Call for Evidence.*

[21] Commodity Futures Trading Commission (CFTC). 2014. Adoption of Rule 575 ("Disruptive Practices Prohibited"). *CME Submission No. 14 – 367.*

[22] Conrad, J., S. Wahal and J. Xiang. 2015. High-frequency quoting, trading, and the efficiency of prices. *Journal of Financial Economics 116(2)*, 271–291.

[23] Cont, R., A. Kukanov and S. Stoikov. 2014. The price impact of order book events. *Journal of Financial Econometrics 12*, 47–88.

[24] Du, S. and H. Zhu. 2014. Welfare and Optimal Trading Frequency in Dynamic Double Auctions. *Working paper*, National Bureau of Economic Research.

[25] Duffie, D. and H. Zhu. 2015. Size discovery. *NBER Working Paper.*

[26] Dufour, A. and R. F. Engle. 2000. Time and the price impact of a trade. *Journal of Finance 55*, 2467–2498.

[27] Egginton, J. F., B. F. Van Ness and R. A. Van Ness. 2014. Quote stuffing. *Louisiana Tech University Working Paper.*

[28] Engle, R. F. and A. Lunde. 2003. Trades and quotes: A bivariate point process. *Journal of Financial Econometrics 1*, 159–188.

[29] European Parliament. 2011. Directive of the European Parliament and of the Council on markets in financial instruments repealing Directive 2004/39/EC of the European Parliament and of the Council, *COM(2011) 656 final.*

[30] Foucault, T., O. Kadan and F. Kandel. 2013. Liquidity Cycles and Make/Take Fees in Electronic Markets, *Journal of Finance 68*, 299–341.

[31] Foucault, T., J. Hombert and I. Roşu. 2016. News Trading and Speed. *Journal of Finance 71*, 335–381.

[32] Friederich, S. and R. Payne. 2015. Order-to-trade ratios and market liquidity. *Journal of Banking & Finance 50*, 214–223.

[33] Garman, M. B. 1976. Market microstructure. *Journal of Financial Economics 3*(3), 257–275.

[34] Goettler, L. R., C. A. Parlour and U. Rajan. 2005. Equilibrium in a Dynamic Limit Order Market, *Journal of Finance 60*, 2149–2192.

[35] Goettler, L. R., C. A. Parlour and U. Rajan. 2009. Informed traders and limit order markets, *Journal of Financial Economics 93*, 67–87.

[36] Gomber, P., M. Haferkorn, and K. Zimmermann. 2016. Securities transaction tax and market quality – the case of France, *European Financial Management 22(2)*, 313–337.

[37] Hagströmer, B. and L. L. Nordén. 2013. The diversity of high-frequency traders. *Journal of Financial Markets 16*, 741–770.

[38] Harris, L. 2002. Trading and exchanges: Market microstructure for practitioners. *Oxford University Press.*

[39] Harris, L. 2013. What to do about high-frequency trading. *Financial Analysts Journal 69(2)*, 6–9.

[40] Hasbrouck, J. 1991. Measuring the information content of stock trades. *Journal of Finance 46*, 179–207.

[41] Hasbrouck, J. and G. Saar. 2013. Low-Latency Trading. *Journal of Financial Markets 16*, 646–679.

[42] Hasbrouck, J. 2015. High Frequency Quoting: Short-Term Volatility in Bids and Offers. *Working paper, SSRN.*

[43] Hautsch, N. and R. Huang. 2012. The market impact of a limit order. *Journal of Economic Dynamics and Control 36*, 501522.

[44] Hendershott, T. J., Jones, C. M. and Menkveld, A. J. 2010. Does Algorithmic Trading Improve Liquidity? *Journal of Finance 66*, 1–33.

[45] Ho, T. and H. R. Stoll. 1981. Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial economics 9 (1)*, 47–73.

[46] Hoffmann, P. 2014. A dynamic limit order market with fast and slow traders. *Journal of Financial Economics 113*, 156–169.

[47] Hollifield, B., R. A. Miller, P. Sandås and J. Slive. 2006. Estimating the Gains from Trade in Limit Order Markets. *Journal of Finance 61*, 2753–2804.

[48] Hunsader, E. S. 2011. Interview: Improving Academic Research into HFT & Fighting High Frequency Spam. *High Frequency Trading Review.*

[49] Jones, C. M. 2013. What do we know about high-frequency trading? *Columbia Business School Research Paper.*

[50] Jovanovic, B. and A. J. Menkveld. 2011. Middlemen in Limit-Order Markets *Working paper*, New York University.

[51] D'Antona Jr., J. 2010. Minimum Quote Life Faces Hurdles, *Traders Magazine.*

[52] Kirilenko, A. A., A. Kyle, M. Samadi and T. Tuzun. 2014. The Flash Crash: The Impact of High Frequency Trading on an Electronic Market, *CFTC Working paper.*

[53] Kirilenko, A. A. and G. Lamacie. 2015. Latency and Asset Prices. *Working paper SSRN 2546567.*

[54] Knorr, E. M. and R. T. Ng. 1997. A unified notion of outliers: Properties and computation. *KDD Conference*, 219–222.

[55] Kyle, A. S. 1985. Continuous auctions and insider trading. *Econometrica 53*, 1315–1335.

[56] Lebedev, N. N., R. A. Silverman and D. Livhtenberg. 1972. Special functions and their applications. *Physics Today 18*, 70.

[57] Lee, C. and M. Ready. 1991. Inferring trade direction from intraday data. *Journal of Finance 46*, 733–746.

[58] Malinova, K. and A. Park. 2015. Subsidizing Liquidity: The Impact of Make/Take Fees on Market Quality, *Journal of Finance 70*, 509–536.

[59] Malinova, K., A. Park and R. Riordan. 2013. Do retail traders suffer from high frequency traders? *Working paper.*

[60] Maskin, E. and J. Tirole. 2001. Markov perfect equilibrium: I. Observable actions. *Journal of Economic Theory 100*, 191–219.

[61] Menkveld, A. J. 2013. High Frequency Trading and The New-Market Makers. *Journal of Financial Markets 16*, 712–740.

[62] Menkveld, A. J. and M. A. Zoican. 2017. Need for speed? Exchange latency and liquidity. *Review of Financial Studies 30*, 1188–1228.

[63] Meyer, S., M. Wagener and C. Weinhardt. 2015. Politically motivated taxes in financial markets: The case of the French financial transaction tax. *Journal of Financial Services Research 47*, 177–202.

[64] Moallemi, C. C. and M. Sağlam. 2013. The Cost of Latency in High-Frequency Trading. *Working paper*, Columbia University.

[65] O'Donoghue, S. M. 2015. The Effect of Maker-Taker Fees on Investor Order Choice and Execution Quality in US Stock Markets. *Working paper.*

[66] O'Hara, M. 2015. High frequency market microstructure. *Journal of Financial Economics 116*, 257–270.

[67] Pakes, A. and P. McGuire. 2001. Stochastic Algorithms, Symmetric Markov Perfect Equilibrium, and the 'curse' of Dimensionality, *Econometrica 69*, 1261–1281.

[68] Pham, M., H. N. Duong and P. Lajbcygier. 2015. Measuring immediate price impact using non-parametric models. *Monash University Working Paper.*

[69] Riordan, R. and A. Storkenmaier. 2012. Latency, liquidity and price discovery. *Journal of Financial Markets 15*, 416–437.

[70] Rojcek, J. and A. Ziegler. 2015. High-frequency trading in limit order markets: Equilibrium impact and regulation. *Swiss Finance Institute Research Paper No. 15-23.*

[71] Rostek, M. and M. Weretka. 2015. Dynamic thin markets. *Review of Financial Studies 28 (10).*, 2946–2992.

[72] Roşu, I. 2014 Fast and slow informed trading. *AFA 2013 San Diego Meetings Paper.*

[73] Securities and Exchange Commission (SEC). 2010. Concept release on equity market structure, *Release No. 34-61358; File No. S7-02-10*, Concept release.

[74] Securities and Exchange Commission (SEC). 2011a. Notice of Filing and Immediate Effectiveness of a Proposed Rule Change Adopting the Text of Financial Industry Regulatory Authority Rule 5210, Which Prohibits the Publication of Manipulative or Deceptive Quotations or Transactions, as NYSE Rule 5210, *Release No. 34-65954; File No. SR-NYSE-2011-61*, Self-Regulatory Organizations; New York Stock Exchange LLC.

[75] Securities and Exchange Commission (SEC). 2011b. Notice of Filing and Immediate Effectiveness of Proposed Rule Change Adopting the Text of Financial Industry Regulatory Authority Rule 5210, Which Prohibits the Publication of Manipulative or Deceptive Quotations or Transactions, as NYSE Arca Equities Rule 5210, *Release No. 34-65955; File No. SR-NYSEARCA-2011-90*, Self-Regulatory Organizations; NYSE Arca, Inc.

[76] Securities and Exchange Commission (SEC). 2012. Notice of Filing and Immediate Effectiveness of Proposed Rule Change to Modify its Excess Order Fee, *Release No. 34-67292; File No. SR-NASDAQ-2012-073*, Self-Regulatory Organizations; NASDAQ Stock Market LLC.

[77] U. S. Securities and exchange commission. 2014. Pilot Plan to Assess Stock Market Tick Size Impact for Smaller Companies, *Release 2014-176.*

[78] Tong, L. 2015. A Blessing or a Curse? The Impact of High Frequency Trading on Institutional Investors. *Working paper.*

[79] Umlauf, S. R. 1993. Transaction taxes and the behavior of the Swedish stock market. *Journal of Financial Economics 33*, 227–240.

[80] Vayanos, D. 1999. Strategic trading and welfare in a dynamic market. *The Review of Economic Studies 66 (2)*, 219–254.

[81] Vives, X. 2011. Strategic supply function competition with private information. *Econometrica 79(6)*, 1919–1966.

[82] Vlachos, M., K. L. Wu, S. K. Chen and S. Y. Philip. 2008. Correlating burst events on streaming stock market data. *Data Mining and Knowledge Discovery 16*, 109–133.

[83] White, H. 2014. *Asymptotic Theory for Econometricians.* London, UK: Academic press.

[84] Zhang, M. Y., J. R. Russell and R. S. Tsay. 2001. A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics 104,* 179–207.